



Genome-Wide Association Study for Identification and Validation of Novel SNP Markers for *Sr6* Stem Rust Resistance Gene in Bread Wheat

Amira M. I. Mourad^{1,2*}, Ahmed Sallam^{1,3}, Vikas Belamkar¹, Stephen Wegulo⁴, Robert Bowden⁵, Yue Jin⁶, Ezzat Mahdy², Bahy Bakheit², Atif A. El-Wafaa², Jesse Poland⁷ and Peter S. Baenziger¹

¹ Department of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, NE, United States, ² Department of Agronomy, Faculty of Agriculture, Assiut University, Assiut, Egypt, ³ Department of Genetics, Faculty of Agriculture, Assiut University, Assiut, Egypt, ⁴ Department of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE, United States, ⁵ United States Department of Agriculture–Agricultural Research Service, Manhattan, KS, United States, ⁶ Cereal Disease Laboratory, United States Department of Agriculture–Agricultural Research Service, St. Paul, MN, United States, ⁷ Plant Sciences Center, Department of Agronomy, Kansas State University, Manhattan, KS, United States

OPEN ACCESS

Edited by:

Hikmet Budak,
Montana State University,
United States

Reviewed by:

Umesh K. Reddy,
West Virginia State University,
United States
Mikhail P. Ponomarenko,
Institute of Cytology and Genetics
(RAS), Russia
Mahmut Tör,
University of Worcester,
United Kingdom

*Correspondence:

Amira M. I. Mourad
amira_mourad@aun.edu.eg

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 12 November 2017

Accepted: 08 March 2018

Published: 27 March 2018

Citation:

Mourad AMI, Sallam A, Belamkar V,
Wegulo S, Bowden R, Jin Y,
Mahdy E, Bakheit B, El-Wafaa AA,
Poland J and Baenziger PS (2018)
Genome-Wide Association Study
for Identification and Validation
of Novel SNP Markers for *Sr6* Stem
Rust Resistance Gene in Bread
Wheat. *Front. Plant Sci.* 9:380.
doi: 10.3389/fpls.2018.00380

Stem rust (caused by *Puccinia graminis* f. sp. *tritici* Erikss. & E. Henn.), is a major disease in wheat (*Triticum aestivum* L.). However, in recent years it occurs rarely in Nebraska due to weather and the effective selection and gene pyramiding of resistance genes. To understand the genetic basis of stem rust resistance in Nebraska winter wheat, we applied genome-wide association study (GWAS) on a set of 270 winter wheat genotypes (A-set). Genotyping was carried out using genotyping-by-sequencing and ~35,000 high-quality SNPs were identified. The tested genotypes were evaluated for their resistance to the common stem rust race in Nebraska (QFCSC) in two replications. Marker-trait association identified 32 SNP markers, which were significantly (Bonferroni corrected $P < 0.05$) associated with the resistance on chromosome 2D. The chromosomal location of the significant SNPs (chromosome 2D) matched the location of *Sr6* gene which was expected in these genotypes based on pedigree information. A highly significant linkage disequilibrium (LD, r^2) was found between the significant SNPs and the specific SSR marker for the *Sr6* gene (*Xcfd43*). This suggests the significant SNP markers are tagging *Sr6* gene. Out of the 32 significant SNPs, eight SNPs were in six genes that are annotated as being linked to disease resistance in the IWGSC RefSeq v1.0. The 32 significant SNP markers were located in nine haplotype blocks. All the 32 significant SNPs were validated in a set of 60 different genotypes (V-set) using single marker analysis. SNP markers identified in this study can be used in marker-assisted selection, genomic selection, and to develop KASP (Kompetitive Allele Specific PCR) marker for the *Sr6* gene.

Highlights:

Novel SNPs for *Sr6* gene, an important stem rust resistant gene, were identified and validated in this study. These SNPs can be used to improve stem rust resistance in wheat.

Keywords: genome-wide association study, linkage disequilibrium, marker-assisted selection, single marker analysis, SNP validation, haplotypes

INTRODUCTION

Stem rust (caused by *Puccinia graminis* f. sp. *tritici* Erikss. & E. Henn.) is one of the most damaging diseases in wheat (Singh et al., 2006). It occurred frequently in the United States from the 1920s to 1960s and caused up to 50% yield losses (Leonard and Szabo, 2005). In the central United States Great Plains, an area from central Texas through central Nebraska, stem rust was a major disease and caused significant reductions in the wheat grain yield (Eversmeyer and Kramer, 2000). However, in recent years, stem rust occurs rarely at Nebraska, in part due to the effective selection for stem rust resistance by the USDA-ARS and University of Nebraska–Lincoln (UNL) Wheat Improvement Team for the past 60 years (Baenziger et al., 2001). The selection was phenotypically performed with little additional information on the genes existing in the germplasm and controlling the resistance except for pedigree information and the use of different pathotypes. Identifying genes controlling stem rust resistance by molecular markers is useful to accelerate breeding programs to improve stem rust resistance by identifying the genes, their vulnerabilities, and being able to pyramid multiple genes in wheat (Nisha et al., 2015).

With the advances in sequencing technologies, genotyping methods that combine variant discovery and genotyping in a single step are now being routinely used in plant breeding research. Genotyping-by-sequencing (GBS) is one such technique that involves the use of restriction enzymes for targeted complexity reduction of the genome followed by multiplexing and sequencing. Genotyping-by-sequencing can generate numerous SNP markers covering a high percentage of the genome in a cost-effective manner (Elshire et al., 2011; Poland and Rife, 2012). Therefore, these genome-wide SNPs can be used in genomic selection, genome-wide association study (GWAS) and genetic diversity studies.

Association mapping (AM) is a powerful approach which identifies polymorphisms near or within a gene of interest that controls the phenotypic differences between genotypes (Soto-Cerda and Cloutier, 2012). To perform AM, it is recommended that 100–500 individuals and codominant SSR or SNP markers are used (Kumar et al., 2011). Allele frequency distribution affects the power of AM to detect an association at the functional polymorphism level. Therefore, rare alleles, present in only a few individuals, cause a problem in association analysis due to their influence on the resolution power of mapping (Pearson and Manolio, 2008). In order to remove the effect of the rare alleles in AM, GBS derived SNPs can be filtered for at least 5% minor allele frequency (Soto-Cerda and Cloutier, 2012).

Few specific SNPs have been published or used for marker-assisted selection (MAS) for stem rust genes. Some exceptions are *Sr2* (Mago et al., 2011), *SrCad* (Kassa et al., 2016) and *Sr36*¹ that have a SNP marker identified and used in MAS. No SNPs have been identified to be associated with some important stem rust resistance genes such as *Sr6*, *Sr30*, *Sr38*, and *Sr24*. Linkage disequilibrium (LD) analysis between specific SSR markers previously identified for stem rust resistance genes and

SNPs generated from GBS with known chromosomal positions will be useful in identifying SNPs that are tightly associated with the stem rust gene of interest. Of course, these linkages between the detected SNPs and the target gene in diverse backgrounds will need to be confirmed.

The objectives of this study were to (1) screen a nursery of 270 Nebraska winter wheat genotypes for their resistance to stem rust race (QFCSC), the common race in the United States, (2) identify SNP markers associated with stem rust resistance using GWAS, (3) validate the SNPs associated with the resistance in another Nebraska winter wheat nursery, and (4) determine whether the markers identified in this study are located in genes and examine their potential role in disease resistance using functional annotations.

MATERIALS AND METHODS

Plant Materials

To identify the SNP markers associated with stem rust resistance, a set of 270 winter wheat genotypes from the 2015 F_{3;6} nurseries (Nebraska Duplicate Nursery – DUP2015, the preliminary yield trial) were used. These genotypes were derived from 800 to 1000 crosses made among primarily Great Plains adapted genotypes with a heavy emphasis on using lines adapted specifically to Nebraska. These 270 genotypes were named as the association set (A-set). To validate the SNP markers associated with resistance in the A-set, a set of 60 genotypes from the 2015 F_{3;7} nurseries (Nebraska Triplicate Nursery–TRP2015, the advanced yield trial) were used and named the validation set (V-set). The TRP2015 nursery is derived from the selections from the DUP2014 nursery and do not overlap with the DUP2015. The V-set genotypes are selected and advanced from the DUP2014 nursery based on the grain yield, grain volume weight, disease resistance, plant height and maturity criteria (El-Basyoni et al., 2013). Three check cultivars were included in the evaluation: ‘Robidoux’ (moderately resistant to moderately susceptible to stem rust race QFCSC), ‘Freeman’ (moderately resistant to moderately susceptible to stem rust race QFCSC), and ‘Goodstreak’ (moderately resistant to stem rust race QFCSC).

Stem Rust Inoculation and Screening

Both sets (A-set and V-set) were evaluated for stem rust resistance using the common stem rust race in Nebraska “QFCSC.” The A-set was evaluated for its resistance in two replications; one at the plant pathology greenhouses, University of Nebraska Lincoln, UNL and the other at USDA-ARS at Kansas State University (KSU). The V-set was evaluated for its resistance in two replications also, one at the plant pathology greenhouses, University of Nebraska Lincoln, UNL and the other at the USDA Cereal Disease Laboratory, St. Paul, MN, United States. The inoculation was performed at the seedling stage as described by Jin and Singh (2006) with minor modifications. Capsules of stored urediniospore were first removed from the freezer and thawed for 30 min at room temperature. Then, they were mixed with lightweight soltrol oil. Primary leaves of 10-day old seedlings were inoculated by atomizing a urediniospores suspension until

¹<http://maswheat.ucdavis.edu/protocols/sr36/>

the leaves were completely covered. Inoculated seedlings were placed in a transparent plastic mist chamber for 12 h in the dark at 20°C. Consequently, seedlings were placed on the greenhouse bench at $20 \pm 2^\circ\text{C}$ and a 16-h photoperiod for 2 weeks. The resistance was scored based on the scale of 0–4 as described by Stakman et al. (1962). Plants with an infection type (IT) score 0–2 were considered as resistant, and plants with IT scores 3–4 were considered as susceptible.

Statistical Analysis of Stem Rust Resistance

Phenotypic data were converted from the Stakman et al. (1962) scale to a linear scale (L-IT) as mentioned in Kumssa et al. (2015) and Zhang et al. (2011) in order to convert the phenotypic data from a qualitative to a quantitative scale for analyses. The converted scores of stem rust resistance ranged from 0 (resistant) to 9 (susceptible). In the converted scale, scores of resistant genotypes ranged from 0 to 5 and those of susceptible genotypes ranged from 6 to 9. For complex IT range scores such as “; 1, 2”, only the lowest (“;”) and highest (“2”) IT scores were converted to the linear scale and then averaged. In both nurseries, the analysis of variance was performed with R software (R Core Team, 2017) using the following model

$$Y_{ij} = \mu + g_i + r_j + e_{ij}$$

where Y_{ij} is an observation of genotype i in replication j , μ is the general mean; g_i and r_j are the main effects of genotypes (fixed effects) and replications (random effects), respectively; e_{ij} is the error. The broad sense heritability (H^2) was calculated using the following formula:

$$H^2 = \sigma_G^2 / (\sigma_G^2 + \frac{\sigma_R^2}{r})$$

where σ_G^2 and σ_R^2 are the variance of the lines and the residuals and r is the number of replicates within the experiment.

DNA Extraction

DNA was extracted from lines in the A-set and V-set for GBS using BioSprint 96 DNA Plant Kits (Qiagen, Hombrechtikon, Switzerland) from 2 to 3 leaves of 2-weeks-old seedlings. For the SSR marker test, DNA from a bulk of six leaves of 5 days old plants was extracted using DNAzol Reagent (Molecular Research Center, Inc. Technical Bulletin 6). The tissue was ground using liquid nitrogen then 300 μl DNAzol reagent was added to this powder and left for 5 min at the room temperature. A volume of 300 μl chloroform was added to the previous mix and left at the room temperature for 5 min before it was centrifuged using Fisher Scientific accuSpin Micor 17 centrifuge (Loughborough, United Kingdom) at $12000 \times g$ for 10 min. The washing process was done in three steps by adding three different washing solutions as follows: (1) absolute alcohol, (2) DNAzol + 75% Ethanol and (3) 75% alcohol only. All genotypes were centrifuged using the Fisher Scientific accuSpin Micor 17 centrifuge for 4 min at $5000 \times g$ after each washing step. The extracted DNA was then re-suspended in 150 μl of TE buffer. DNA concentration was measured

using spectrophotometry (Gen5™ microplate reader and image software with Take3™ micro-volume plates [BioTek, Winooski, VT, United States]).

Genotype-by-Sequencing (GBS)

The DNA of both sets (A-set and V-set) were genotyped using GBS by digesting the DNA with two restriction enzymes, *PstI* and *MspI* (Poland and Rife, 2012). Pooled libraries were sequenced using Illumina, Inc. NGS platforms. The FASTQ file containing the raw data of sequence reads were processed for SNP identification using TASSEL 5.0 v2 GBS pipeline (Bradbury et al., 2007). Chinese Spring genome assembly from the International Wheat Genome Sequencing Consortium (IWGSC) Reference Sequence v1.0 was used as the reference genome. Raw sequence data of the tested genotypes along with additional ~3000 breeding lines from the University of Nebraska wheat program were combined to increase the genome coverage and read depth for SNP discovery in both nurseries. SNP markers identified were filtered for minor allele frequency (MAF > 0.05), maximum missing sites per SNP < 20% and maximum missing sites per genotype < 20% (Belamkar et al., 2016). Heterozygous loci were then marked as missing to obtain better estimates of marker effects (Peter Bradbury, personal communication) and the SNP markers were re-filtered using the same filtering criteria.

Population Structure

The analysis of population structure was performed on the A-set using the Bayesian model-based software program STRUCTURE 3.4 (Pritchard et al., 2000). For each run, burn-in iteration was 100,000, followed by 100,000 Markov chain Monte Carlo (MCMC) replications after burn-in. The admixture and allele frequencies correlated models were considered in the analysis. Five impeded iterations were used in the STRUCTURE. The hypothetical number of the subpopulation (k) extended from 1 to 10. The best k was identified according to Evanno et al. (2005) using STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

Genome-Wide Association Study (GWAS)

Stem rust resistance data of the A-set were used to identify SNPs associated with the resistance. The association of the SNP markers retained after quality-filtering and stem rust measurements was carried out using TASSEL 5.0 software (Bradbury et al., 2007) using a mixed linear model (MLM:

TABLE 1 | Analysis of variance for stem rust resistance in the association set (A-set) of 270 genotypes and the validation set (V-set) of 60 genotypes.

Source	A-set		V-set	
	D.F.	M.S.	D.F.	M.S.
Replications	1	20.15*	1	23.3
Genotypes	269	6.75**	59	13.40**
Error	263	4.08	58	6.07
Total	533		118	
Broad-sense heritability (%)	39.59		54.68	

* $p < 0.05$, ** $p < 0.01$.

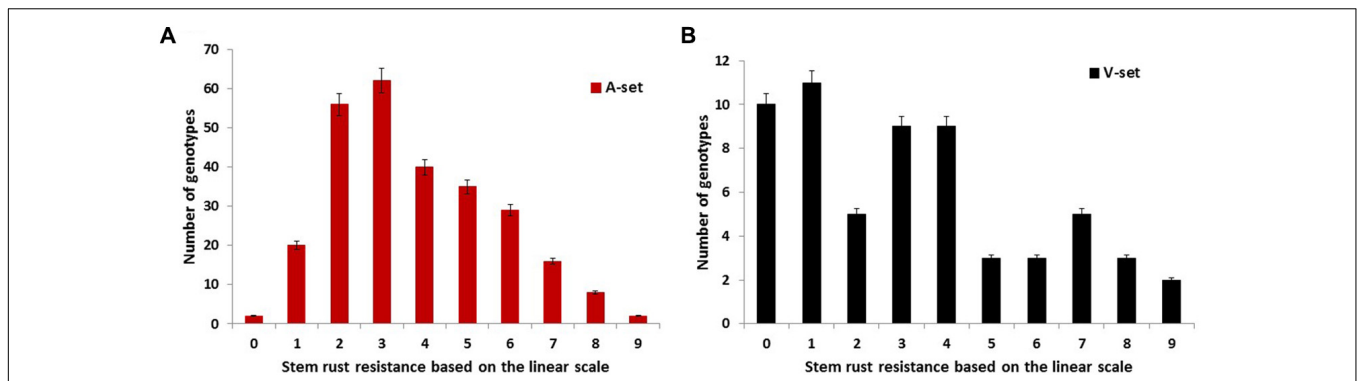


FIGURE 1 | The infection response of the genotypes in the (A) association set (A-set, 270 genotypes) and (B) validation set (V-set, 60 genotypes) to infection with the common stem rust race in Nebraska, QFCSC using the linear scale of Kumssa et al. (2015).

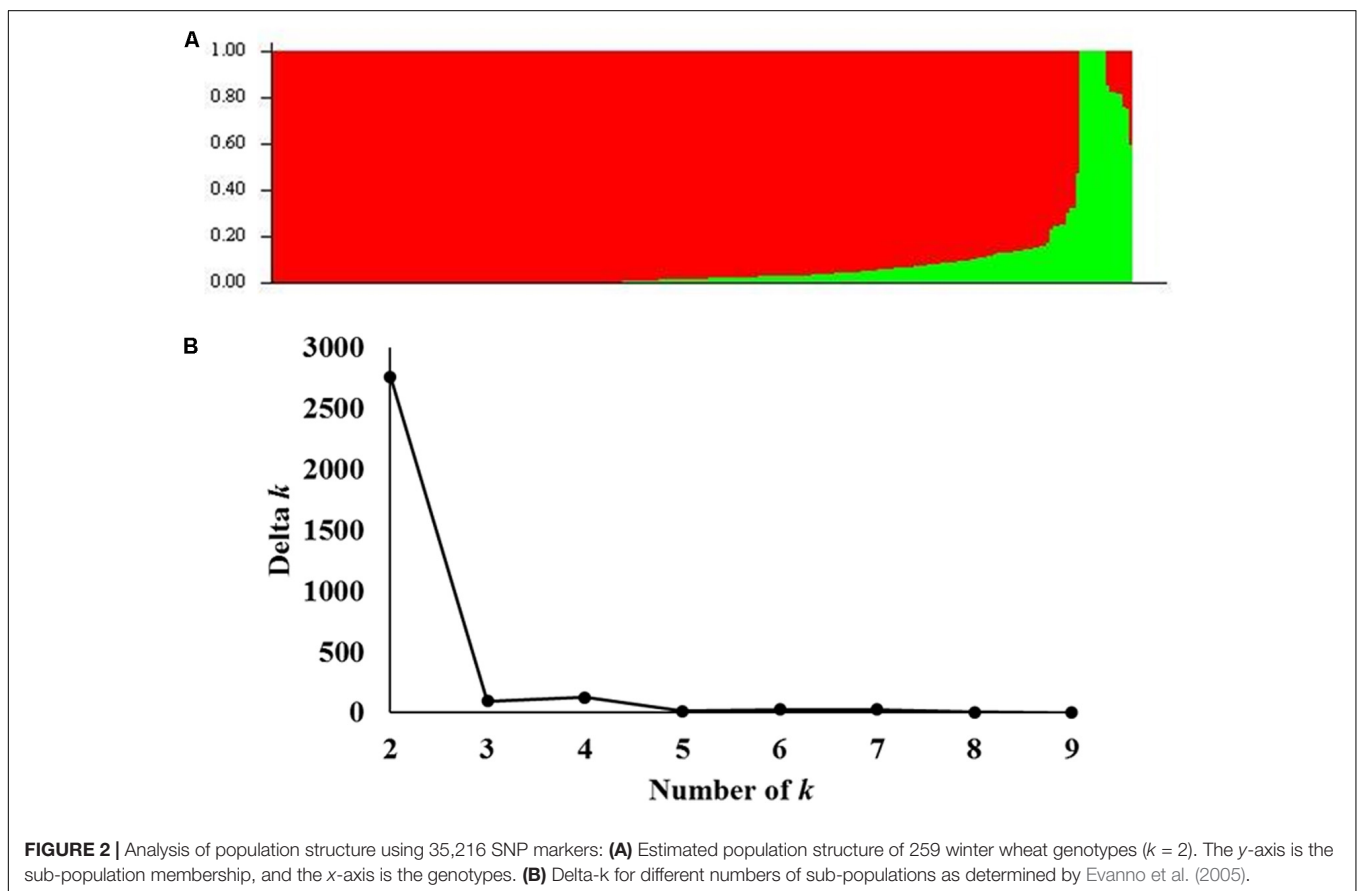


FIGURE 2 | Analysis of population structure using 35,216 SNP markers: (A) Estimated population structure of 259 winter wheat genotypes ($k = 2$). The y-axis is the sub-population membership, and the x-axis is the genotypes. (B) Delta-k for different numbers of sub-populations as determined by Evanno et al. (2005).

Yu et al., 2006). The marker-trait association was tested against Bonferroni corrections at a significance level of 5%. The allele estimates of each marker determine the influence of the allele on the phenotype. In the linear scale used for phenotypic measurements, smaller values represent resistance and larger values indicate susceptibility. Hence, alleles with a lower marker effect score are effective alleles and are linked to increasing resistance in the population. The phenotypic variation explained by a marker (R^2) was calculated for the significant SNPs using TASSEL 5.0 (Bradbury et al., 2007). The Q-Q and Manhattan

plots of the GWAS were developed using 'qqman' R package (Turner, unpublished).

For the significant SNP markers detected by GWAS, the alleles associated with increased stem rust resistance were assigned a value of 1 while those associated with decreased stem resistance were marked 0 to perform a correlation analysis with the specific *Sr6* SSR marker. Linkage disequilibrium (r^2) among the significant SNPs and the specific SSR marker located on the same chromosome was calculated by TASSEL 5.0 and visualized as a heatmap using 'LDheatmap' R package (Shin et al., 2006).

TABLE 2 | Association analysis of stem rust resistance using the mixed linear model (MLM) and the correlation between the significant SNPs and *Xcfd43* markers which tags *Sr6*.

SNP ID	Chro.	P-value	Target allele ⁽¹⁾	Allele effect ⁽²⁾	R ² ⁽³⁾ (%)	Corre.Xcfd43
S2D_57511425	2D	2.49E-09	G:A	-1.85	16.16	0.83
S2D_61759932	2D	3.92E-09	C:T	-1.86	16.75	0.90
S2D_59551557	2D	4.46E-09	G:A	-1.82	16.25	0.90
S2D_59658161	2D	6.27E-09	G:A	-1.83	16.59	0.90
S2D_58726937	2D	6.38E-09	T:C	-1.81	15.96	0.92
S2D_60028884	2D	7.45E-09	A:G	-1.79	16.08	0.88
S2D_60028887	2D	7.45E-09	A:G	-1.79	16.08	0.88
S2D_58806022	2D	9.09E-09	G:C	-1.85	16.6	0.90
S2D_58740240	2D	2.11E-08	G:T	-1.72	14.84	0.92
S2D_58892256	2D	2.23E-08	A:G	-1.76	15.43	0.89
S2D_60546609	2D	2.26E-08	T:C	-1.81	17.68	0.87
S2D_58725975	2D	3.58E-08	T:C	-1.69	14.89	0.91
S2D_58725982	2D	3.58E-08	C:T	-1.69	14.89	0.91
S2D_60546009	2D	3.65E-08	A:G	-1.72	14.74	0.91
S2D_58776075	2D	3.97E-08	C:T	-1.67	13.99	0.91
S2D_58726903	2D	5.81E-08	A:G	-1.7	14.46	0.90
S2D_57322511	2D	7.96E-08	G:C	-1.65	13.82	0.84
S2D_56248408	2D	1.07E-07	C:T	-1.67	12.94	0.91
S2D_58775439	2D	1.08E-07	C:G	-1.62	13.18	0.89
S2D_61284312	2D	1.41E-07	C:A	-1.68	13.52	0.87
S2D_56298113	2D	1.57E-07	A:G	-1.62	13.18	0.90
S2D_59653673	2D	2.15E-07	T:C	-1.6	12.35	0.86
S2D_55920637	2D	2.43E-07	C:G	-1.58	12.75	0.92
S2D_56190285	2D	2.45E-07	C:G	-1.58	12.58	0.86
S2D_56190288	2D	2.45E-07	C:G	-1.58	12.58	0.86
S2D_60963807	2D	2.67E-07	T:C	-1.75	14.13	0.85
S2D_60963818	2D	2.67E-07	A:G	-1.75	14.13	0.85
S2D_57151331	2D	3.38E-07	C:T	-1.56	12.86	0.90
S2D_56445930	2D	4.95E-07	T:A	-1.57	13.54	0.91
S2D_56445931	2D	4.95E-07	A:C	-1.57	13.54	0.91
S2D_57359928	2D	7.50E-07	C:A	-1.49	10.99	0.91
S2D_57151324	2D	9.88E-07	C:A	-1.48	11.7	0.89

⁽¹⁾The allele on the left increased the resistance. ⁽²⁾The effect of left allele associated with increased resistance. ⁽³⁾Phenotypic variation explained by marker. Highlighted SNPs locate in haplotype block 81. In each SNP_ID, S2D refers to chromosome 2D followed by the position of that SNP on the chromosome.

Haplotype Block Analysis

Haplotype block analysis was performed using Haploview 4.2 software (Barrett et al., 2005). The SNP data of the A-set was used to calculate the pair-wise LD between SNPs. A cutoff of 1% was used, meaning that if the addition of a SNP to a block resulted in a recombinant allele at a frequency exceeding 1%, then that SNP was not included in the block. The Haplotype blocks were constructed using the four-gamete methods which create block boundaries where there is evidence of recombination between adjacent SNPs based on the presence of all four gametic types (Wang et al., 2002).

Polymerase Chain Reaction (PCR) Conditions

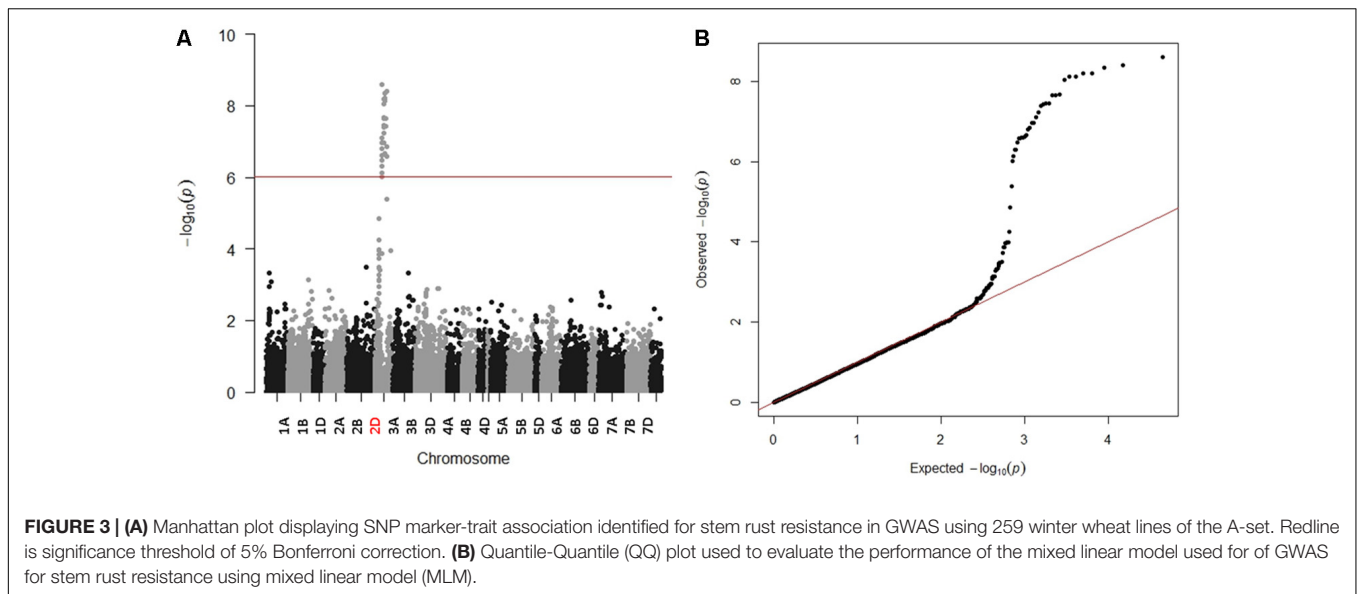
Polymerase chain reaction (PCR) for the available SSR primers of the *Sr6* gene (*Xcfd43* and *Xwmc453*) was performed in 15 μ l volume with 2 μ l 20 ng DNA template, 3 μ l GoTaq flexi buffer

(without MgCl₂), 0.3 μ l 0.25 mM dNTPs, 1.20 μ l 25 mM MgCl₂, 0.2 μ l from 0.5 μ l/ μ l GoTaq Flexi Taq Polymerase (Promega, Madison, Wisconsin, USA) and 0.75 μ l 10 pmol from each primer. SSR marker products were separated in super fine resolution (SFR) 3% Agarose gel. The differential line for the *Sr6* gene (ISr6-Ra CI 14163) was included in the SSR test to identify the target band size for each primer. The SSR products gel was scored visually and by using Gel Analyzer 2010 software.²

Single Marker Analysis (SMA) for Marker Validation

Phenotypic data of the V-set, using the converted linear scale (Kumssa et al., 2015), as well as genotypic data of markers which were associated with stem rust in the A-set were used to perform single marker analysis. The analysis was done by using

²<http://www.gelanalyzer.com/>



SAS Version 9 (SAS Institute Inc., 2013) following this model:

$$Y = \mu + f(\text{marker}) + \text{error},$$

where Y is equal to the trait value, μ is equal to the population mean, and $f(\text{marker})$ is a function of the significant markers (Francis et al., 2011).

Genes Containing Significant SNPs and Their Annotations

To further investigate the GWAS results, we inspected whether any of the significant SNPs were in genes identified and annotated in the reference genome assembly (IWGSC Ref Seq v1.0). The effect of the significant SNPs on annotated genes was determined by using SnpEff (Cingolani et al., 2012). Functional annotation of the genes harboring significant SNPs were retrieved from the genome annotations provided by IWGSC and examined for their association with stem disease resistance.

RESULTS AND DISCUSSION

The analysis of variance revealed highly significant differences among genotypes (Table 1), indicating our phenotypic assay was successful. Such a high genetic variation found among genotypes could be very useful for plant breeders to select the most resistant genotypes to stem rust. Figure 1A represents the response of the A-set to the most common Nebraska stem rust race (QFCSC). The A-set was mostly resistant (80% of the genotypes) with L-IT ranging from 0 to 5. This result was expected because the previous generations were screened for stem rust resistance and the selection was made using stem rust resistance as one of the selection criteria.

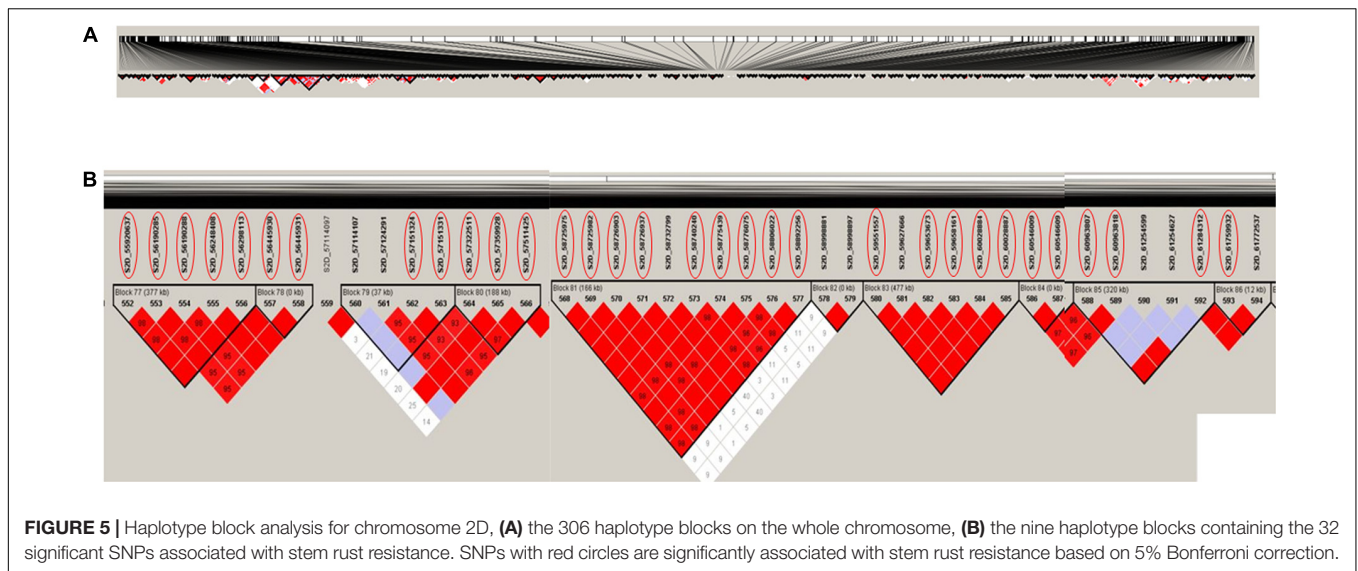
Association Mapping for Stem Rust Resistance Population Structure

The GBS generated a set of 35,216 SNPs after filtering for minor allele frequency (MAF > 0.05), maximum missing sites per SNP < 20% and maximum missing sites per genotype < 20% (Belamkar et al., 2016) (Data are available on doi: 10.6084/m9.figshare.5948416). This set of SNPs was used in population structure analysis. After marking the heterozygous loci as missing values and re-filtering the data, 23,053 SNPs and a set of 259 genotypes in the A-set nursery and 60 genotypes in the V-set were used in our marker-trait association studies. The 23,051 SNPs were used in the GWAS analysis.

Population structure analysis was performed on the A-set and two possible subpopulations were identified (Figure 2A). To confirm this result, the number of suggested k was plotted against the calculated Δk obtained from STRUCTURE software. A clear peak was observed for $k = 2$ (Figure 2B). Therefore, a k value of two was chosen to describe the genetic structure of the genotypes used in this study. That the structure was found in the A-set was expected due to the selection process for important agronomic traits and different climatic zones in the Nebraska wheat breeding program.

Genome-Wide Association Study and Linkage Disequilibrium

Due to the presence of structure in our tested population, MLM was used for GWAS. Association analysis identified 32 significant SNPs for stem rust resistance based on Bonferroni correction ($\alpha = 0.05$) (Table 2 and Figures 3A,B). Remarkably, all the significant SNPs were located on chromosome 2D. The phenotypic variation explained by each SNP marker (R^2) ranged from 11.70 to 17.68%. The R^2 for all significant SNPs indicates that these SNPs represented a major QTL



Several previously identified stem rust resistance genes such as *Sr6*, *Sr46* and *Sr32* are mapped to chromosome 2D.³ However, based on the available pedigree information of the tested populations, we expected to have the *Sr6* gene and not to have *Sr32* and *Sr46* genes in our materials. To confirm the presence of the *Sr6* gene in the tested population, specific SSR markers (*Xcfd43* and *Xwmc453*) were used in genotyping the genotypes. Only the *Xcfd43* marker presented clear amplification, while *Xwmc453* did not give good amplification. The result of the *Xcfd43* marker revealed that of the 270 genotypes, 123 (47%) were found to be containing the target allele associated with the *Sr6* gene. The correlation between the specific SSR marker and each one of the significant SNPs is presented in **Table 2**. The 32 significant SNPs revealed high correlations with the specific SSR marker with r^2 values ranging from 0.83 for marker S2D_57511425 to 0.92 for markers S2D_58726937, S2D_58740240 and S2D_55920637. This result indicates that these SNPs were tightly linked to *Sr6*. A High LD was found between the SSR marker (*Xcfd43*) and all the significant SNPs (**Figure 4**). This high LD between the specific SSR marker for the *Sr6* gene and the significant SNPs suggested (1) the SSR marker is highly co-inherited with the 32 SNPs, and (2) it is considered as a further support that some of these significant SNPs could be a part of this gene. Such high LD is very useful for marker-assisted selection. The significant SNPs were removed from the marker data and the AM was re-analyzed to find out if the LD among the 32 markers interferes with the effect of other SNPs with small effects and associated with stem rust resistance. No significant markers were associated with stem rust resistance. This result is additional evidence of the power of GWAS performed in this study.

To confirm that the significant 32 SNPs are identifying the *Sr6* gene, we inspected the location of the *Sr6* SSR primer using

the IWGSC RefSeq v1.0 marker mapping file⁴. The mapping file released by the IWGSC includes location information for previously mapped SSR markers. The sequence of the SSR marker (*Xcfd43*) has been mapped to a scaffold that is currently not included in any of the chromosomes and is placed along with other scaffolds that lack location information in an “Unknown” chromosome in the IWGSC RefSeq v1. Due to the high LD between the *Sr6* SSR marker and the significant SNPs as well as the previous study that has placed the SSR marker for *Sr6* on chromosome 2D (Tsilo et al., 2010), we believe the scaffold containing the *Sr6* sequence can be placed on chromosome 2D.

Genes Underlying Significant SNPs and Their Functional Annotations

To further understand the association between the significant SNPs and the stem rust resistance, we reviewed the annotation of genes containing these SNPs and studied the effect of the SNPs on the genes. Out of the 32 SNPs, eight SNPs are located within genes: TraesCS2D01G108000.1 (two SNPs), TraesCS2D01G104700.1 (two SNPs), TraesCS2D01G104400.1 (one SNP), TraesCS2D01G104600.1 (one SNP), TraesCS2D01G106100.1 (one SNP) and TraesCS2D01G107200.1 (one SNP). Four of these eight SNPs result in an amino acid change (missense variant), one of the SNPs causes the loss of stop codon, and the rest of the three are synonymous SNPs (Supplementary Table 1). The SNP in the gene, TraesCS2D01G106100.1, results in loss of stop codon and is the highest impact variant. The loss of stop codon results in degradation of the transcript in the nucleus, and thus a complete loss of function of the gene (Fasken and Corbett, 2005). The functional annotation of some of the genes indicates their potential involvement in fungal disease resistance (**Table 3**). For example, TraesCS2D01G104700.1 and TraesCS2D01G104600.1

³<http://maswheat.ucdavis.edu/protocols/StemRust/StemRustMap.html>

⁴<https://wheat.pw.usda.gov/cgi-bin/graingenes/report.cgi?class=probe;name=CFD43>

TABLE 4 | Single marker analysis (SMA) for the 32 significant SNPs from the association set (A-set) and the *Xcfd43* marker which predicts the presence of *Sr6* in the validation set (V-set).

SNP ID	P-value	F-value	(R ²) (%)	Corr. between SNP and SSR marker
S2D_55920637	0.0037	9.24**	15.09	0.40
S2D_56190285	0.0014	11.57**	20.10	0.46
S2D_56190288	0.0014	11.57**	20.10	0.46
S2D_56248408	<0.0001	26.94**	35.47	0.64
S2D_56298113	<0.0001	25.60**	32.99	0.57
S2D_56445930	<0.0001	27.00**	34.62	0.60
S2D_56445931	<0.0001	27.00**	34.62	0.60
S2D_57151324	<0.0001	27.00**	34.62	0.60
S2D_57151331	<0.0001	27.00**	34.62	0.60
S2D_57322511	<0.0001	19.55**	28.52	0.61
S2D_57359928	<0.0001	26.30**	34.03	0.60
S2D_57511425	0.0002	16.62**	25.33	0.56
S2D_58725975	<0.0001	22.90**	30.99	0.57
S2D_58725982	<0.0001	22.90**	30.99	0.57
S2D_58726903	<0.0001	26.60**	35.18	0.63
S2D_58726937	<0.0001	30.96**	38.72	0.64
S2D_58740240	0.0002	16.12**	23.66	0.50
S2D_58775439	<0.0001	25.60**	32.99	0.57
S2D_58776075	<0.0001	30.21**	37.20	0.65
S2D_58806022	<0.0001	30.52**	37.91	0.66
S2D_58892256	<0.0001	27.90**	36.28	0.59
S2D_59551557	<0.0001	26.60**	35.18	0.63
S2D_59653673	0.0002	16.46**	25.54	0.56
S2D_59658161	<0.0001	29.63**	36.30	0.62
S2D_60028884	<0.0001	22.90**	30.99	0.57
S2D_60028887	<0.0001	22.90**	30.99	0.57
S2D_60546009	<0.0001	24.77**	32.69	0.57
S2D_60546609	<0.0001	23.64**	36.57	0.56
S2D_60963807	0.0048	8.91**	18.22	0.51
S2D_60963818	0.0048	8.91**	18.22	0.51
S2D_61284312	<0.0001	26.63**	35.21	0.56
S2D_61759932	<0.0001	25.60**	32.99	0.57
<i>Xcfd43</i>	<0.0001	20.90**	27.18	—

R² the phenotypic variation explained by the marker. **Highly significant association between the SNP marker and the stem rust resistance in the V-set.

are WRKY transcription factors and the role of the WRKY transcription factor family is well-known in biotic stresses including disease resistance (Pandey and Somssich, 2009; Wang et al., 2013; Phukan et al., 2016). The TraesCS2D01G104400.1 is an E3-ubiquitin-protein ligase and these genes have been shown to contribute to resistance to fungal diseases such as powdery mildew (caused by *Blumeria graminis* f. Sp. *tritici*) (Zhu et al., 2015). The F-box domain containing protein (TraesCS2D01G106100.1) was found to have an effect on disease resistance in tomato (*Solanum lycopersicum* L.) and tobacco (*Nicotiana tabacum* L.) (Maclean et al., 2008). In summary, many of the genes underlying significant SNPs have been annotated as being involved in disease resistance which further validated the marker-trait associations identified in this study.

Haplotype Block Analysis

Haplotype block analysis revealed the existence of 306 haplotype blocks on chromosome 2D (Figure 5A). The 32 significant SNPs were located within nine haplotype blocks with numbers 77, 78, 79, 80, 81, 83, 84, 85, and 86 (Figure 5B). Block 81 contains most of the significant SNPs (nine SNPs), while the remaining markers were distributed in eight different blocks (Figure 5B). All the 32 significant SNPs are located on haplotype blocks adjacent to each other (on block number 77 through 86). This result indicated that these SNPs most likely identify the same QTL. The haplotype blocks shed light on genomic regions associated with the trait following a GWAS study. For example, selecting the haplotype block enriched with significant SNPs as compared to using individual SNPs can reduce redundancy introduced by using all SNPs (Mokry et al., 2014). Based on the GWAS and haplotype block analysis, we recommend the significant SNPs located in block 81 (Table 2) as the best targets for marker-assisted selection to improve stem rust resistance in seedling winter wheat.

Validation of the SNPs in the Validation Set (V-set)

To validate the association between the significant SNPs and stem rust resistance, a set of 60 genotypes was evaluated for QFCSC stem rust race in two replications. Highly significant differences were found between the genotypes with no significant differences between the replications (Table 1). Seventy-eight percent of the genotypes were resistant with L-IT scores ranging from 0 to 5 which indicated the possibility of these genotypes to be used for the validation purpose (Figure 1B).

Based on single marker analysis, all the significant SNPs identified in the A-set had a strong significant association with stem rust resistance in the V-set (Table 4). The *F-values* of the significant SNPs ranged from 8.91 to 30.96. The *R²* for the significant SNPs ranged from 15.09 to 38.72% which indicates that these SNPs explain the high percentage of the phenotypic variation in the V-set.

To confirm that the 32 SNPs were good markers for the *Sr6* gene, the V-set was genotyped using specific SSR marker for the *Sr6* gene (*Xcfd43*). Thirty-seven percent (22 genotypes) had the target band associated with *Sr6*. The correlation between the SSR marker and the significant SNPs (Table 4) ranged from 0.40 to 0.64. The LD between the SSR and the significant SNPs (Figure 6) was low. This result was explained by the high percentage of the missing SNPs in the V-set with a range extending from 6.78 to 28.81% (Figure 7). The high percentage of missing loci reduced the correlation and LD values. The significant association between the 32 SNPs and stem rust resistance in the V-set is evidence that these SNPs are very useful for a future breeding program to improve stem rust resistance in winter wheat where Ug99 pathotypes are not present. Validation of QTLs associated with target traits is one of the main steps of marker-assisted selection. The advantage of the V-set is that it represents a different genetic background from that used for GWAS (A-set). Generally, the 32 SNPs associated with stem rust resistance especially those that are located in block 81

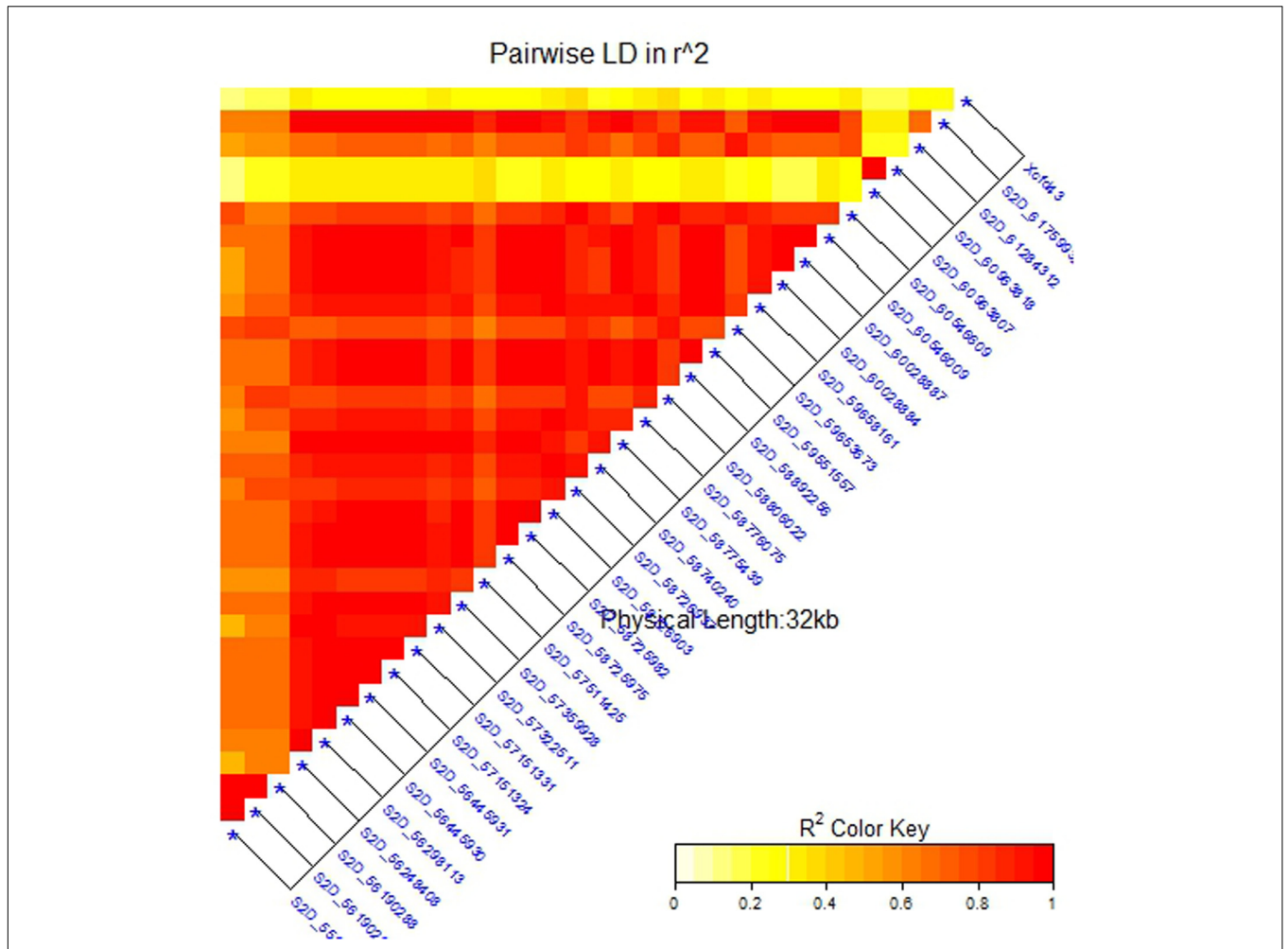


FIGURE 6 | Linkage disequilibrium (LD) analysis in the validation set (V-set): heatmap of LD between the significant SNPs detected by the mixed linear model (MLM) (at significance level of 5% Bonferroni correction) and SSR marker *Xcfd43* that predicts *Sr6*.

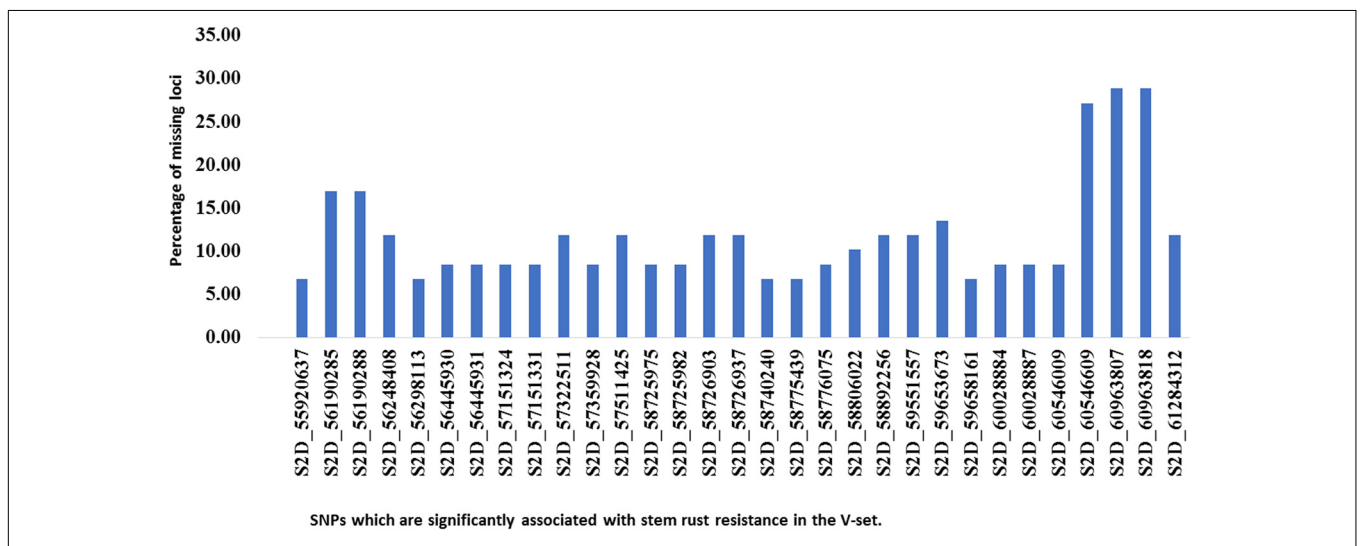


FIGURE 7 | Histogram of the percentage of missing loci for each significant SNP (identified in the association set) in the validation set.

(nine SNPs) could be converted to Kompetitive Allele Specific PCR (KASP) for the *Sr6* gene.

CONCLUSION

In conclusion, our study for the first time identified 32 novel SNPs associated with *Sr6* which is an important gene providing resistance to a wide range of stem rust races (MCCFC, QCCSM, QFCSC, RCRSC, GFMNC, and TPMKC). The SNPs associated with stem rust in this study can be considered for MAS. However, the nine SNPs located within the same haplotype block and explain relatively high phenotypic variation are probably most promising. The reasons supporting this conclusion are (I) the high LD and correlation found between these SNPs and the *Xcfd43* marker (predicts the *Sr6* gene) and (II) the validation of the association between these SNPs and the *Sr6* gene with highly significant p and R^2 values in a set of lines with different genetic backgrounds. These novel SNPs should be very useful for improving stem rust resistance.

AUTHOR CONTRIBUTIONS

AM performed all the genetic and phenotyping analyses and wrote the manuscript. AS helped in genetic and phenotypic analyses. VB performed the SNP calling from GBS data and helped in the genetic analysis. SW, RB, and YJ helped in phenotyping stem rust resistance. EM, AE-W, and BB helped in design the study. JP performed GBS of the two populations. PB designed the study, helped in discussing the results and drafted

the manuscript. All the authors agreed to be accountable for the content of the work.

FUNDING

This work was financially partial supported by cultural affairs and mission sector, the Egyptian government.

ACKNOWLEDGMENTS

The authors would like to thank Raghuprakash Kastoori Ramamurthy, a Postdoctoral Research Associate at the University of Nebraska–Lincoln, for his positive support during the lab work and Javed Sidiqi, a Ph.D. student at the University of Nebraska–Lincoln, for his support during stem rust inoculation at UNL. Population structure analysis was completed utilizing the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative. The authors would also like to thank Natasha Pavlovikj from the University of Nebraska–Lincoln, Holland Computing Center for her help in data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00380/full#supplementary-material>

REFERENCES

- Baenziger, P. S., Shelton, D. R., Shipman, M. J., and Graybosch, R. A. (2001). Breeding for end-use quality: reflections on the Nebraska experience. *Euphytica* 119, 95–100. doi: 10.1023/A:1017583514424
- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Belamkar, V., Guttieri, M. J., El-Basyoni, I., Hussain, W., Poland, J., Jarquin, D., et al. (2016). *Integration of Genomic Selection in the Nebraska Wheat Breeding Program*. Available at: <https://pag.confex.com/pag/xxiv/webprogram/Paper21286.html> [accessed January 25, 2018].
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, snpEFF: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Earl, D. A., and vonHoldt, B. M. (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
- El-Basyoni, I., Baenziger, P. S., Dweikat, I., Wang, D., Eskridge, K., and Saadalla, M. (2013). Using DArT markers to monitor genetic diversity throughout selection: a case study in Nebraska's winter wheat breeding nurseries. *Crop Sci.* 53, 2363–2373. doi: 10.2135/cropsci2013.01.0051
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Evanno, G., Regnaut, S., and Goudet, J. (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
- Eversmeyer, M. G., and Kramer, C. L. (2000). Epidemiology of wheat leaf and stem rust in the central great plains of the USA. *Annu. Rev. Phytopathol.* 38, 491–513. doi: 10.1146/annurev.phyto.38.1.491
- Fasken, M. B., and Corbett, A. H. (2005). Process or perish: quality control in mRNA biogenesis. *Nat. Struct. Mol. Biol.* 12, 482–488. doi: 10.1038/nsmb945
- Francis, D. M., Merk, H. L., and Namuth-covert, D. (2011). *Introduction to Single Marker Analysis (SMA)*. 3–5. Available at: <http://www.extension.org/pages/32552/introduction-to-single-marker-analysis-sma>
- SAS Institute Inc (2013). *Base SAS 9.4 Procedures Guide Statistical Proceed*, 2nd Edn. Cary, NC: SAS Institute Inc.
- Jin, Y., and Singh, R. P. (2006). Resistance in U.S. wheat to recent Eastern African isolates of *Puccinia graminis* f. sp. *tritici* with virulence to resistance Gene *Sr31*. *Plant Dis.* 90, 476–480. doi: 10.1094/PD-90-0476
- Kassa, M. T., You, F. M., Fetch, T. G., Fobert, P., Sharpe, A., Pozniak, C. J., et al. (2016). Genetic mapping of SrCad and SNP marker development for marker-assisted selection of Ug99 stem rust resistance in wheat. *Theor. Appl. Genet.* 129, 1373–1382. doi: 10.1007/s00122-016-2709-z
- Kumar, J., Pretep, A., Solanki, R., Gupta, D., Goyal, A., and Chaturvedl, S. (2011). Advances in genomics resources for improving food legume crops. *J. Agric. Sci.* 150, 289–318. doi: 10.3389/fpls.2015.00563
- Kumssa, T. T., Baenziger, P. S., Rouse, M. N., Guttieri, M., Dweikat, I., Brown-Guedira, G., et al. (2015). Characterization of stem rust resistance in wheat cultivar gage. *Crop Sci.* 55, 229–239. doi: 10.2135/cropsci2014.05.0348

- Leonard, K. J., and Szabo, L. J. (2005). Stem rust of small grains and grasses caused by *Puccinia graminis*. *Mol. Plant Pathol.* 6, 99–111. doi: 10.1111/J.1364-3703.2004.00273.X
- Macleod, D., Takken, F. L. W., and Jones, J. D. G. (2008). The F-box protein ACRE189 / ACIF1 regulates cell death and defense responses activated during pathogen recognition in tobacco and tomato. *Plant Cell* 20, 697–719. doi: 10.1105/tpc.107.056978
- Mago, R., Brown-Guedira, G., Dreisigacker, S., Breen, J., Jin, Y., Singh, R., et al. (2011). An accurate DNA marker assay for stem rust resistance gene Sr2 in wheat. *Theor. Appl. Genet.* 122, 735–744. doi: 10.1007/s00122-010-1482-7
- Mokry, F. B., Buzanskas, M. E., Mudadu, M. D. A., Grossi, A., Higa, R. H., Ventura, R. V., et al. (2014). *Linkage Disequilibrium and Haplotype Block Structure in a Composite Beef Cattle Breed*. Available at: <http://www.biomedcentral.com/1471-2164/15/S7/S6>
- Nisha, R., Sivasamy, M., Gajalakshmi, K., Shajitha, P., Vikas, V. K., Peter, J., et al. (2015). Pyramiding of stem rust resistance genes to develop durable and multiple disease resistant wheat varieties through marker aided selection. *Int. J. Extensive Res.* 5, 1–9.
- Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990
- Pearson, T. A., and Manolio, T. A. (2008). How to interpret a genome-wide association study. *JAMA* 299, 1335–1344. doi: 10.1001/jama.299.11.1335
- Phukan, U. J., Jeena, G. S., and Shukla, R. K. (2016). WRKY transcription factors: molecular regulation and stress responses in plants. *Front. Plant Sci.* 7:760. doi: 10.3389/fpls.2016.00760
- Poland, J. A., and Rife, T. W. (2012). Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1111/j.1471-8286.2007.01758.x
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing.
- Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of Pairwise linkage Disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* 16, 1–10.
- Singh, R., Hodson, D. P., Jin, Y., Huerta-Espino, J., Kinyua, M., Wanyera, R., et al. (2006). Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 1, 1–13. doi: 10.1079/PAVSNR20061054
- Soto-Cerda, B. J., and Cloutier, S. (2012). “Association mapping in plant genomes,” in *Genetic Diversity Plants*, ed. M. Caliskan (London: InTech), 29–54. doi: 10.5772/2640
- Stakman, E. C., Stewart, D. M., and Loegering, W. Q. (1962). *Identification of Physiologic Races of Puccinia graminis var. Tritici*. Washington, DC: United States Government Printing Office.
- Tsilo, T. J., Jin, Y., and Anderson, J. A. (2010). Identification of flanking markers for the stem rust resistance gene Sr6 in wheat. *Crop Sci.* 50, 1967–1970. doi: 10.2135/cropsci2009.11.0648
- Wang, C., Deng, P., Chen, L., Wang, X., Ma, H., Hu, W., et al. (2013). A wheat WRKY transcription factor TaWRKY10 confers tolerance to multiple abiotic stresses in transgenic tobacco. *PLoS One* 8:e65120. doi: 10.1371/journal.pone.0065120
- Wang, N., Akey, J. M., Zhang, K., Chakraborty, R., and Jin, L. (2002). Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.* 71, 1227–1234.
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., 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
- Zhang, D., Bowden, R., and Bai, G. (2011). “A method to linearize stakman infection type ratings for statistical analysis,” in *Proceedings of the 2011 Borlaug Global Rust Initiative Technical Workshop*, St. Paul, MN, 17–20.
- Zhu, Y., Li, Y., Fei, F., Wang, Z., Wang, W., Cao, A., et al. (2015). E3 ubiquitin ligase gene CMPG1 – V from *Haynaldia villosa* L. contributes to powdery mildew resistance in common wheat (*Triticum aestivum* L.). *Plant J.* 84, 154–168. doi: 10.1111/tj.12966

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Mourad, Sallam, Belamkar, Wegulo, Bowden, Jin, Mahdy, Bakheit, El-Wafaa, Poland and Baenziger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.