Pak. J. Bot., 54(4): 1405-1416, 2022. DOI: http://dx.doi.org/10.30848/PJB2022-4(11)

GENOME-WIDE ASSOCIATION MAPPING AND POPULATION STRUCTURE FOR STRIPE RUST IN PAKISTANI WHEAT GERMPLASM

RIZWAN QAISER[1*], ZAHID AKRAM[1*], SHAHZAD ASAD[2], INAM-UL-HAQ[3], SAAD IMRAN MALIK[1], MUHAMMAD FAYYAZ[2], MUHAMMAD SUFIYAN[4], SAHIR HAMEED KHATTAK[5], KARANSHER SINGH SANDHU[6] AND GAGANJOT SINGH SIDHU[6]

[1]Department of Plant Breeding and Genetics, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan
[2]Crop Disease Research Institute, National Agriculture Research Centre, Islamabad, Pakistan
[3]Deptt.of Plant Pathology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan
[4]Crop Diseases Research Institute, Sunny Bank, Murree, Pakistan
[5]National Institute for Genomics & Advanced Biotechnology, National Agricultural Research Centre, Islamabad, Pakistan
[6]Department of Plant Breeding and Genetics, Washington state University, USA.
[*]Corresponding author's email: cmspak@ymail.com & akranzahid@hotmail.com

Abstract

Stripe rust disease caused by *Puccinia striiformis* is one of the utmost destructive foliar diseases of wheat all around the world. The most effective strategy to control this disease is to develop and grow resistant varieties. To identify the genomic regions responsible for resistance, 294 Pakistani hexaploid wheat accessions were subjected to genome-wide association studies (GWAS). These lines were characterized phenotypically for stripe rust response at seedling stage in controlled conditions in Pakistan and in fields near Mount Vernon and Pullman in Washington, United States. A targeted amplicon sequencing approach was used to genotype the wheat germplasm with 787 of SNP markers. Twenty-four genotypes showed resistance to stripe rust in controlled conditions in Pakistan. In Washington, 193 and 97 genotypes showed resistance at Mount Vernon and Pullman, respectively. GWAS results showed that seven and three loci were associated with resistance observed at the seedling stage under controlled and field conditions, respectively, and only one locus on chromosome 7A was significantly associated with adult plant resistance. This study identified resistance loci in Pakistani wheat germplasm that can be used in breeding resistant wheat varieties.

Key words: Stripe rust, *Triticum aestivum*, GWAS, Infection, and adult-plant resistance.

Introduction

Wheat is an extensively cultivated crop all over the world, especially in Asia. Worldwide nearly 215m ha of the land is cultivated for hexaploid (*Triticum aestivum* L.) and tetraploid (*T. turgidum* L. subsp. *durum*) wheat. A total of 95 mha that comprise 44% of the world is used for cultivation in Asia. The main contributors are Pakistan, China, and India. China alone holds 62 mha out of 95mha for wheat growers. Wheat is considered a staple food that feeds almost 40% of the total world population (Waqar *et al.,* 2018). The anticipated global wheat requirement for 9.6 billion individuals in 2050 is forecasted to be met by increasing wheat production by 60% from 2000-2050. The world wheat production in 2000 and 2018 was 583 and 733 million metric tons, respectively. Concurrently, wheat suffers from three devastating fungal diseases, including stripe (yellow), stem (black), and leaf (brown) rust, resulting in huge yield lost.

Being the fifth largest country for wheat production 25.7 million metric tons in 2019-2020 (www.fao.org) where this crop is cultivated everywhere in the country. Yellow or stripe rust (produced by Puccinia striiformis), is a globally significant disease of wheat (Waqar *et al.,* 2018). *Puccinia striiformis* Westend. *f.* sp. *tritici* Eriks. (Pst) is known to cause stripe rust disease, which is a significant foliar disease. The severity of stripe rust are visible in Pakistan's northern areas, upland, and foothills of Baluchistan. The disease has historically jeopardized wheat production when epidemics occurred at various times in 1947-48, 1953-54, 1958-59, 1977-78, and 1992-93. Stripe rust decreases the mass and quality of the seed and fodder. Cost is also reduced for seeds yield that is struck by rust as they have low vigor and poor germination rate along with shriveled grains. The losses can go up to or nearly cent percent depending upon the variety whether resistant or not, early, or late maturity rate of infection and development and favorable environment and period duration of rust disease (Chen, 2005). Pakistani wheat is heavily grown but nearly 70% of the total production is affected by rust with decrease yield and cost for farmers.

Long-term fertility building requires a combined methodology instead of a short-range approach and targeted way out instead of conventional agriculture approaches (Rehman *et al*., 2020). Therefore combined efforts of agriculture scientists, especially plant breeders and geneticists are required to overcome loss inflicted by a lethal disease. In the past wheat breeding, has relied on the successful development of rust resistant wheat varieties. However, the continued emergence of new virulent races often shortens the life of the deployed varieties and keeps the breeding for stripe rust resistance a tough challenge and needs long term planning. Genetically, resistance is characterized into seedling resistance and adult plant resistance (APR). The seedling resistance is mostly controlled by few major race specific genes which expresses throughout all growth stages. Most of the varieties containing seedling resistance have become susceptible because new pathogen that is more virulent, are emerging. Conversely, APR is quantitatively controlled by single or multiple genes with relatively slight effects are usually race non-specific and expresses only or more effectively

at the post seedling stage. More than 80 stripe rust resistance genes have been permanently named Yr (yellow rust) genes. There are nearly 300 genes/QTL maps have been reported with provisional names (Wang & Chen, 2017). The majority of the Yr genes are race-specific, but still, many genes are not race-specific, for example, *Yr18, Yr29, Yr30*, and *Yr46*. The mentioned genes unaccompanied are inadequate, but they can provide decent levels of resistance when used in combination. *Yr29* which is closely linked with *Lr46* for leaf rust resistance confers a moderate level of APR to stripe disease (William *et al*., 2003). Merging the all-stage resistance and APR genes in wheat varieties would be the most effective strategy to provide high-level protection and mitigation of the damage caused by new races. Hence, it is important to detect more resistance sources to avoid future stripe-rust epidemics in Pakistan where year-round inoculum is present.

Traditionally, bi-parental QTL mapping has been successfully used to identify linked genetic markers for stripe resistance in genotypes (William *et al*., 2003). In present years, GWAS based on linkage disequilibrium (Goldstein & Weale, 2001) has become popular as it is an alternative method with several improvements by analysis based mapping over normal conventional linkages. Eliminating the need to make crosses and generate mapping populations, GWAS is the most effective approach that utilizes natural variation provides broader allele coverage, and exploits the higher number of meiotic events that happened throughout the evolutionary history of the germplasm. This has permitted in a faster decay of linkage disequilibrium (LD) and mapping of the associated loci with the target traits to a much smaller genomic region RIL (recombinant inbred lines) or DH (doubled haploid) populations (Rafalski, 2010; Nordborg & Weigel, 2008). The term LD is referred to historically as an increase (dis-equilibrium) of specific rust alleles at different loci and the stages of LD extent can be measured statistically. GWAS is also helpful in terms of the possibility of using historically characterizing data phenotypically (Kraakman, 2006). GWAS has been successfully used to identify marker-trait associations for the resistance of stripe rust in wild emmer wheat (Sela *et al*., 2014), synthetic hexaploid wheat (Zegeye, 2014), emmer wheat (Liu *et al*., 2017), winter wheat (Bulli, 2016), spring wheat (Godoy *et al*., 2018; Tadesse *et al*., 2015) and other wheat diseases (Kollers *et al*., 2013). However, the utilization of an association mapping panel to dissect the genetic basis of stripe rust resistance in Pakistani wheat germplasm has not been demonstrated yet.

The current study aimed at conducting GWAS for stripe rust in Pakistani *Triticum aestivum* collection of 294 genotypes comprised of advanced breeding lines, candidate lines, commercial wheat varieties, and local landraces of Pakistan. The objectives of present study were: 1) to evaluate Pakistani wheat germplasm for their resistance to prevailing races of stripe rust across three environments and conditions, 2) to assess the population structure of this germplasm based on SNP markers, and 3) to conduct GWAS for genetic markers associated with loci governing resistance against stripe rust.

Materials and Methods

Germplasm: A hexaploid wheat collection of 294 genotypes obtained from National Agricultural Research Centre (NARC), Islamabad and Plant Genetic Resources Institute (PGRI) were selected for this study. The collection consisted of advanced breeding lines, candidate lines, commercial wheat varieties, and local landraces of Pakistan including 49 National Uniform Wheat Yield Trial lines of Pakistan. Morocco was used as the susceptible spreader of PST inoculum in the study.

Greenhouse and field experiments at Pakistan: Seedling stage experiment was conducted by testing the 294 wheat genotypes with four Pakistani Pst isolates representing four predominant races (Table 1). The seedling test was conducted in the greenhouse at Crop Disease Research Institute, Murree. The experiment was conducted in augmented design with replicates performed in the year 2016-17. For each genotype, 10-15 seeds were planted in one pot containing Sunshine mix #1 growing medium placed in plastic trays (5 pots/tray). At 12 days after sowing, urediniospores suspended in mineral and petroleum ether (30:70 concentration) were sprayed onto plants and then allowed to air dry, and the inoculated seedlings were incubated in a dew chamber for 48 hours with 100% relative humidity at 9°C with 18 hours daily photoperiod. Afterward, these plant seedlings were transferred to a screen house, with the controlled temperature at 12-18°C. Different types of infection were noted after 15 days of inoculation (Line & Qayoum, 1992).

The wheat panel was also evaluated in fields at Crop Diseases Research Institute, NARC, Islamabad, Pakistan (33°40'13.5"N, 73°07'33.6"E) and Cereal Crops Research Institute (CCRI), Pirsabak, Nowshera, Pakistan (34°01'02.0"N, 72°02'59.3"E). Stripe rust evaluations were conducted under artificial epidemics during the wheat cropping season in the years 2015 and 2016. Thus, four year-location combinations were conducted. Each of the genotypes was planted in two rows adjacent to each other as 1 m long head rows followed by every third row of Morocco following the local recommended agronomic management practices. These rows were planted in augmented design where all the genotypes were distributed in six blocks and each block consisted of 50 genotypes in addition to Morocco. Morocco was also planted in the surroundings for inoculum dispersal for uniform disease development for the study. A spore suspension containing multiple diverse races that were preserved at the CDRI department was sprayed in Morocco. The mixed inoculum had virulence to all resistance genes in the 18 Yr single-gene differentials, except *Yr5, Yr10, Yr15*, and *YrSP* (Wan *et al*., 2016). Inoculations were done twice – first with "mineral oil" along with petroleum ether and second with normal water by addition of some spores and Tween 20 (2-3 drops) of at booting / tillering and stages respectively. Data collection on all the genotypes started as soon as susceptible Morocco developed 60-70 percent rust severity while data was collected twice and the average was used for performing GWAS analysis. IT was scored based on the 0-to-9 scale similar to seedling evaluation as mentioned above, and SEV was documented visually as percent infection (for rust) on the individual plant using a modification of Cobb's gauge (Peterson *et al*., 2011).

Table 1. Avirulence/virulence profiles of Pakistani races of *Puccinia striiformis* f. sp. *tritici* (*Pst*) used in this study.

Pst Isolate	Race[a] Name	Octal code	*Yr* genes Avirulence	Virulence
	PSTv-101	161266	*Yr1,Yr5,Yr9,Yr10,Yr15,Yr24,Yr32,YrSP,Yr76*	*Yr6,Yr7,Yr8,Yr17,Yr27,Yr43,Yr44,YrTr1,YrExp2*
	PSTv-76	571262	*Yr5,Yr10,Yr15,Yr24,Yr32,YrSP,YrTr1,Yr76*	*Yr1,Yr6,Yr7,Yr8,Yr9,Yr17,Yr27,Yr43,Yr44,YrExp2*
	PSTv-220	541760	*Yr5,Yr8,Yr9,Yr10,Yr15,YrSP,YrTr1,YrExp2,Yr76*	*Yr1,Yr6,Yr7,Yr17,Yr24,Yr27,Yr32,Yr43,Yr44*
	PSTv-221	561242	*Yr5,Yr9,Yr10,Yr15,Yr24,Yr32,Yr44,YrSP,YrTr1,Yr76*	*Yr1,Yr6,Yr7,Yr8,Yr17,Yr27,Yr43,YrExp2*

(Add the isolates)

[a] Names and octal codes of *Pst* races followed Wan *et al*. (2016)

Field evaluation in Washington State: The wheat genotypes were also evaluated for stripe rust response in the fields near Pullman (46°45'30.0"N, 117°11'35.4"W) and Mount Vernon (48°26'24.0"N, 122°23'14.9"W), Washington in the year of 2018. The Mount Vernon site was planted on April 26 and the Pullman site on May 10. For each entry, about 5-gram seeds were grown in almost 50 cm row and spaced nearly 20 cm apart, with the susceptible check Avocet S (AvS) planted every 20 rows and surrounding all nurseries. Stripe rust IT and Sev data were collected on June 7 and June 27 when plants were at middle jointing (Zadoks GS 31) and flowering (Zadoks GS 60) (Zadoks *et al*., 1974), respectively at Mount Vernon. At Pullman, stripe rust IT and SEV data were recorded on July 17, when plants were flowering (Zadoks GS 60).

Targeted amplicon sequencing for SNP genotyping: Genomic DNA from each genotype was extracted from fresh leaf tissues using the Cetyl Trimethyl Ammonium Bromide (CTAB) method (Doyle & Doyle, 1987). Extracted DNA was subjected to targeted amplicon sequencing (TAS) as mentioned by Bernardo *et al*., (2015) along with slight modifications as described in Pupo *et al*., (2019). Two rounds of PCR were performed to amplify the target regions. In the first round of PCR, 1K locus-specific primers were used and in the second round of PCR, the normal Ion An adapter followed by a unique barcode specific for each genotype was added at the 5' end of each amplicon. These amplicons were then purified using the QIA quick PCR purification kit (Qiagen, Germany) and filtered using the Agencourt AM Pure XP beads (Beckman Coulter, USA). Following this, two size selections were performed. In the first selection, fragments ranging from 100–350 base-pair length were selected, extracted from a 4% Size Select E-Gel (Life Technologies, USA), and purified using the QIA quick Gel Extraction Kit. In the second selection, fragments of length ~300 base pairs were selected on a 2% Size Select E-Gel, purified using the QIA quick purification kit, and quantified through the Qubit dsDNA HS assay kit (Life Technologies, USA). Sequencing was performed on an Ion Torrent Proton Sequencer (Thermo Fisher Scientific, USA) at the USDA-ARS Genotyping Laboratory in Pullman, WA. The sequenced data was done as mentioned by an in-house pipeline (Skinner *et al*. unpublished) and a total of 940 SNP markers were obtained. To identify the polymorphic markers, filtering was done to remove monomorphic markers, markers with a minor allele frequency (MAF) <0.05, and markers with 20% or more missing data. Filtered SNPs were then used in statistical analysis for GWAS.

Population linkage and structure dis-equilibrium estimation: The population structure based on filtered SNP markers was analyzed in STRUCTURE V2.3.3 (Pritchard *et al*., 2000). To assign the individuals into sub-populations (K), the admixture model of population structure was applied. Following an initial burn-in of 20,000 iterations, the number of hypothetical subpopulations (K) was set from 1 to 10, with 50,000 Monte Carlo Markov Chain (MCMC) replicates. For each K, five independent runs were performed, and the most optimal number of subpopulations (K) were determined using the method elaborated by Evanno *et al*., (2005). The linkage disequilibrium (LD) was observed employing the filtered loci in TASSEL 5.0 software (Bradbury *et al*., 2007). The LD decay rate was analyzed individually for each chromosome.

ANOVA and heritability: Analysis of variance (ANOVA) was conducted on infection types and severity using the PROC MIXED procedure including genotype, location, year, genotype by location interaction as random effects. The overall average was considered as a fixed effect. Different variance components were computed using the REML method. For each location, the model used is listed below:

$$Yij = Gi + Ej + GEij + eij$$

where Gi the effect of the genotype, Ej is the year effect, GEij is the interaction between the genotype and year, and eij is the residual. Across all locations, the following model was used.

$$Yijl = Gi + EKjl + GEKijl + eijl$$

where Gi the effect of the genotype, EKjl is the effect of the interaction of year and location, GEKijl is the interaction between the genotype and year and location, and eijl is the residual. The broad-sense heritability (H2) estimates for each trait was calculated for each location and across all locations as.

$$H2 = (2G)/(2G + (2E/y) + (2GE/y) + (e2/y))$$

"where 2G is the genotypic variance, 2E is the environmental variance, 2GE is the genotype * environment variance and e2 is the residual variance and y corresponds to the number of years for each location and the number of years multiplied by the number of locations across locations". Descriptive statistics and Pearson correlation coefficients for locations and years were calculated for IT and SEV values using the SAS PROC UNIVARIATE and PROC CORR procedures. The BLUPs for IT and SEV across locations and years were computed using the PROC MIXED.

GWAS analyses for stripe rust resistance: GWAS for loci associate with IT and SEV response at seedling and adult plant stages were conducted on 287 genotypes using a total of 787 high-quality SNP markers. Different general linear (GLM) and mixed linear (MLM) association models were compared to select the best model for marker-trait association using the GAPIT and Farm CPU software implemented in R (Liu *et al*., 2016). They tested six models included a fixed general linear model containing kinship only (K GLM), a fixed general linear model containing kinship and population structure using the first three principal components (PC3) (K+Q GLM), a compressed mixed linear model containing kinship only (K CMLM), a compressed mixed linear model containing kinship and population structure using the first three principal components (K+Q CMLM), and two Farm CPU models– one without correction for population structure and one containing PC3 as a correction for population structure (Lipka *et al*., 2012). The population structure (Q matrix) was determined through principal component analysis and family relatedness (K matrix) using the method of Van Raden in GAPIT (Vanraden, 2008). The Q and K matrices were fitted as fixed and random effects into the model to ensure the detection of only genetically significant associations. Models were compared based on the deviation of observed probability from expected distribution in the Q-Q plot. Association analysis was carried out by estimating the marker-wise threshold ($p \leq 0.005$), based on the Bonferroni correction ($\alpha = 0.10$), and the p-value threshold recommended by Farm CPU for each trait for each environment (location year) separately and the BLUPs across environments for the adult-plant responses under both artificially inoculated and natural infections.

Results

Phenotypic responses: A total of 294 genotypes were evaluated for IT to four Pakistani Pst races at the seedling stage. The avirulence /virulence formulae of the four races are provided in Table 1. The IT responses of the genotypes varied greatly among Pst races (Fig. 1). The widest range (1-9) was observed among genotypes for the response to race PSTv-101 whereas relatively narrow ranges were observed in tests with races PSTv-220 (2-8) and PSTv-221 (1-7). The genotypes showed the highest median IT response (6.5) for PSTv-76 whereas the lowest median response (4.0) was observed for Pst-IV. For artificial epidemics in Pakistan environments, the lowest IT score of 2 and a median score of 6 was observed across all environments (Fig. 2A; Table 2). The distribution of genotype responses was almost similar (C2015, C2016, N2015) except for N2016. The distribution of genotype responses for SEV was more variable than IT under artificial epidemics (Fig. 2B). C2015, N2015, and N2016 showed a similar median SEV score of 40 while a median score of 30 was observed for C2016 (Table 2). Combining the data over two years for each of the locations indicated the NARC site to be fairly more symmetrical than the CCRI location for IT based on the computed skewness whereas CCRI showed more symmetrical distribution for SEV (Table 2). Under natural conditions in Washington State, the U.S.A, a median IT score of 2 was observed for both the Mount Vernon and Pullman locations despite the wide range (2-8) of IT distribution, whereas the IT variation was wider for the Pullman location (Fig. 2C). For SEV scores under natural conditions, similar distributions were found at the two locations although Mount Vernon showed a higher median SEV score (15%) than Pullman (5%) (Fig. 2D). However, both locations showed a highly skewed distribution (1.22 – 2.21) for both phenotype traits (Table 3). With an exception of IT at Pullman, the computed kurtosis indicated tailed distributions for IT and SEV under the Washington natural disease conditions (Table 3).

The percentages of wheat genotypes with resistance to races PSTv-101, PSTv-76, and PSTv-220 were relatively low (22-25%) compared to race PSTv-221 (42%) (Fig. 3). The IT scores for PSTv-101, PSTv-76, and PSTv-221 showed a bimodal distribution. For PSTv-220, the IT scores had a close to normal distribution. The genotypes that showed susceptibility to PSTv-101, PSTv-76, PSTv-220, and PSTv-221 were 49%, 50%, 21%, and 25%, respectively. The stripe rust responses of all genotypes are provided in (**Table 1)**. Further, 24 genotypes were found resistant to all-important races (four) at the seedling stage (**Table 2 and Supplementary Fig. 1).** When genotypes were tested under artificially inoculated field conditions in Pakistan, none of the genotypes revealed resistance to stripe rust in all of the four environments. In the U.S. under natural Pst infection, 97 genotypes were resistant at the middle jointing stage (Zadoks GS 31) whereas 193 genotypes had resistance at the adult-plant stage for both locations, indicating that 96 of the 193 genotypes had adult-plant resistance.

Table 2. Descriptive statistics and estimated heritability of wheat genotypes in different field environments in Pakistan

Statistic	CCRI		NARC		Across environments	
	IT[a]	SEV[b]	IT[a]	SEV[b]	IT[a]	SEV[b]
Number	574	574	574	574	1148	1148
Minimum	2	5	0	0	0	0
Maximum	9	90	9	90	9	90
Mean	6.21	33.28	6.20	39.45	6.20	36.36
Median	6	30	6	40	6	30
St. Dev.	1.28	16.8	1.50	23.54	1.40	20.68
Skewness	-0.10	0.29	0.03	0.43	-0.02	0.54
Kurtosis	-0.22	-0.72	0.17	-0.83	0.11	-0.40
σ^2_g	0.35**	51.22*	1.70**	428.12**	0.74**	177.89**
σ^2_e	0.00	0.00	0.04^{ns}	0.00	0.01^{ns}	12.00^{ns}
σ^2_{ge}	0.69**	230.14**	0.00	125.84**	0.43**	240.01**
σ^2_{res}	0.60	0.99	0.54**	1.00	0.77	0.99
H^2	0.35	0.31	0.85	0.87	0.71	0.74

[a]= IT and [b] = SEV. Asterisks * and ** indicate $p<0.05$ and $p<0.001$

Table 3. Descriptive statistics and estimated heritability of wheat genotypes in different field environments in US.

Statistic	Pullman		Mount Vernon		Across locations	
	IT[a]	SEV[b]	IT[a]	SEV[b]	IT[a]	SEV[b]
Number	282	282	277	277	559	559
Minimum	2	0	2	2	2	0
Maximum	8	100	8	100	8	100
Mean	3.48	18.01	2.89	20.77	3.19	19.38
Median	2	5	2	15	2	10
St. Dev.	2.24	29.23	1.60	18.16	1.97	24.40
Skewness	1.22	2.00	2.08	2.21	1.59	2.10
Kurtosis	-0.10	2.49	3.59	5.38	1.17	3.65
σ^2_g	-	-	-	-	2.47**	406.48**
σ^2_e	-	-	-	-	0.17^{ns}	2.91^{ns}
σ^2_{ge}	-	-	-	-	0.02^{ns}	185.52**
σ^2_{res}	-	-	-	-	1.29	0.99
H^2	-	-	-	-	0.77	0.81

[a] = IT and [b] = SEV

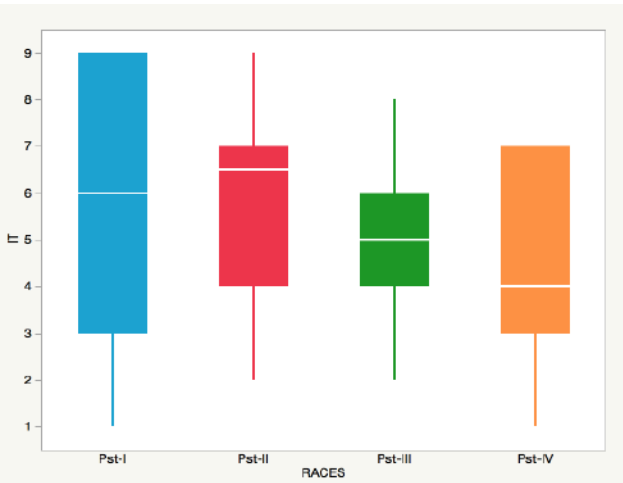


Fig. 1. Box plot distribution of IT in response to four stripe rust pathogen races at seedling stage in controlled conditions.

Table 4. Pearson's correlation coefficients for IT and SEV response of wheat genotypes grown in Pakistan.

Environment	C2015	C2016	N2015	N2016
IT vs. IT				
C2015	1	-	-	-
C2016	0.21**	1	-	-
N2015	0.43**	0.19*	1	-
N2016	0.38**	0.26**	0.76**	1
SEV vs. SEV				
C2015	1	-	-	-
C2016	0.18*	1	-	-
N2015	0.59**	0.16*	1	-
N2016	0.56**	0.19**	0.77**	1
IT↓ vs. SEV→				
C2015	0.58**	0.11^{ns}	0.36**	0.33**
C2016	0.14*	0.60**	0.10^{ns}	0.16**
N2015	0.31**	0.15*	0.66**	0.63**
N2016	0.34**	0.19**	0.55**	0.74**

Asterisks * p<0.05 and ** p<0.001

Genotype adjusted means were also calculated based on BLUPs. The IT and SEV BLUPs computed for Pakistani artificial field conditions were normally distributed. However, the distribution of BLUPs for both IT and SEV under the U.S. natural infections showed a skewed distribution (**Supplementary Fig. 2**) similar to that of combined data (Table 3).

Trait correlation and estimates of heritability: For the field tests in Pakistan under artificial inoculation, the Pearson correlation coefficients (r) stripe rust IT and SEV were highly significant (p<0.001) (Table 4). The highest correlation was found for the growing season of 2015 between NARC and CCRI averaged 0.43 and 0.59 for the IT and SEV data, respectively. Average correlations within locations were higher for the NARC location with correlation coefficients of 0.76 and 0.77 for IT and SEV, respectively. When IT and SEV were compared to each other across different locations and years, correlations at the NARC location (0.55–0.74) were usually higher and more significant than the CCRI location (0.11–0.60). For the tests in the Washington States under natural infections, the correlation coefficients for IT and SEV were also significant (p<0.001) for the Pullman and Mount Vernon locations with r values of 0.69 and 0.77, respectively (Table 5). Correlations between IT and SEV within Pullman and Mount Vernon were relatively higher than the correlation between IT and SEV between the two locations. The respective correlation coefficients within Pullman and Mt. Vernon were 0.87 and 0.91 and between Pullman and Mount Vernon were 0.74 and 0.69.

Estimates of variance components indicated significant (p<0.001 and 0.05) differences for IT and SEV among the genotypes across the CCRI and NARC locations (Table 2). Likewise, significant differences were found for both IT and SEV when ANOVA was performed across all environments. Genotype environment interactions were also significant for both traits, locations, and across environments except IT at the NARC location. Estimation of broad-sense heritability using the REML method indicated high H2 estimates for the NARC location and across environments, ranging from 0.85 to 0.71 for IT and 0.87 to 0.74 for SEV. For the field tests under natural infections in Washington State, only across environments, ANOVA was performed as only one-year data were recorded within each location. Genotypes were highly significant for both IT and SEV whereas genotype * environment interactions were significant only for SEV. As compared to artificial inoculation tests in Pakistan, greater heritability estimates (0.77 and 0.81 for IT and SEV) were observed in the Washington tests when broad-sense heritability was computed across environments (Tables 2 and 3).

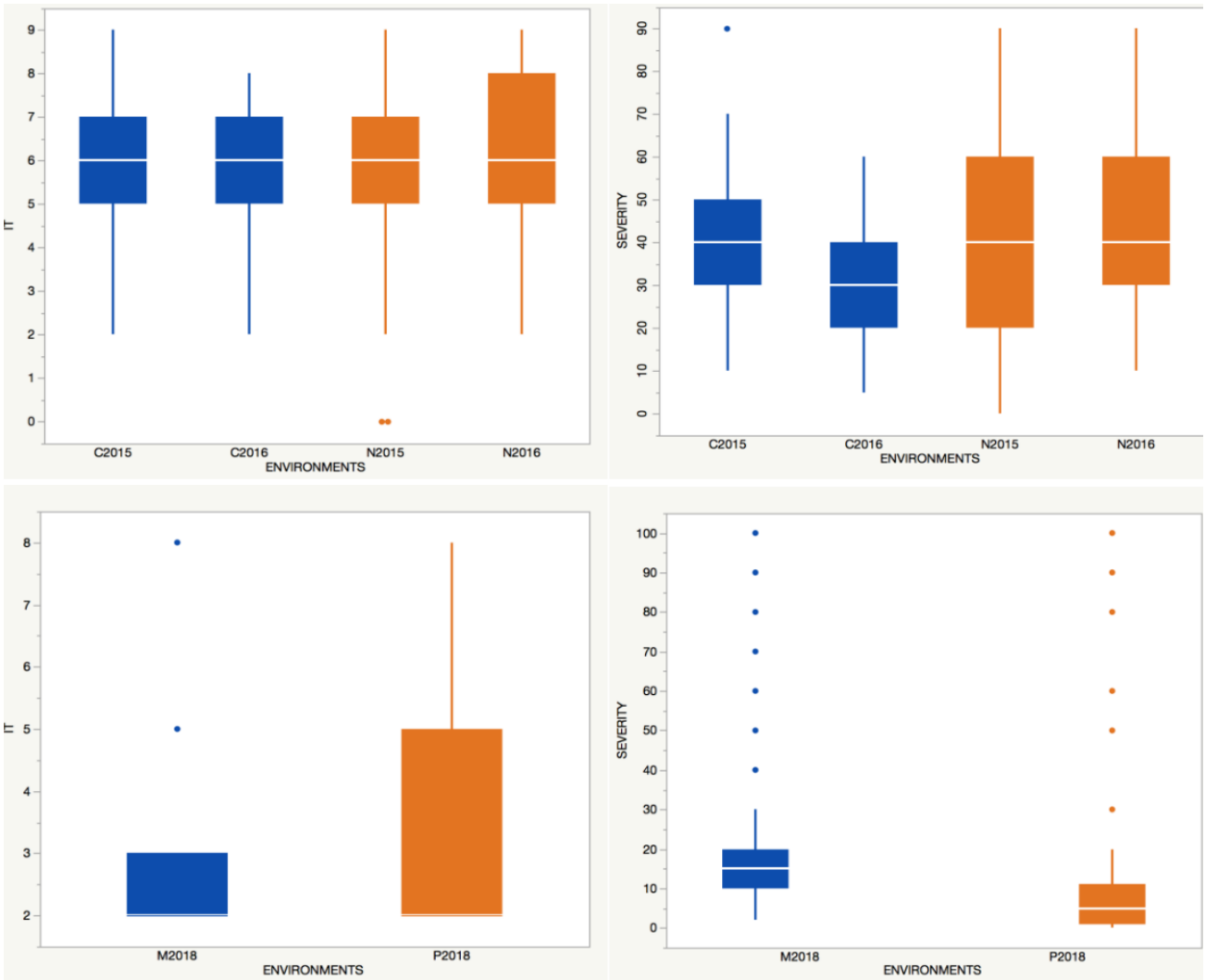


Fig 2. Box plots showing IT and SEV distributions. (A) IT distributions across location × year Pakistan environments under artificial epidemics in CCRI 2015 (C2015), CCRI 2016 (C2016), NARC 2015 (N2015), and NARC 2016 (N2016). (B) SEV distributions across four location×year Pakistan environments. (C) IT distributions across two US locations under natural infections in Mount Vernon 2018 (M2018) and Pullman 2018 (P2018). (D) SEV distributions across two US locations.

Structure of population used in the present study: The analysis was done by using 787 high-quality SNP markers distributed across 21 wheat chromosomes (**Supplementary Fig. 3**) after filtering for MAF > 0.05 and missing data < 20%. According to the STRUCTURE analysis, the 287 genotypes can be subdivided into six subgroups based on the highest k value observed at 6 (Fig. 4A). The clustering of the genotypes into six subgroups is shown in (Fig. 4B). Moderate levels of genetic relatedness among the genotypes were identified as depicted by the heat map of the kinship matrix of the 287 genotypes (Fig. 4C). Principal component analysis (PCA) also grouped population into size groups (supplementary Fig 4). Results clearly depicted broad genetic basae of the population as 6 clear distinct groups were observed.

Association analysis of SNP markers to stripe rust response: Six models comprising the GLM, CMLM, and the FarmCPU model implemented in the R package, were compared by testing the association of the phenotypic traits in the present study. The quantile-quantile (QQ) plot was used to determine the spurious associations. Large deviance of the detected values of P from the expected values of P for the GWAS results under the null hypothesis of no association between SNP markers and the corresponding traits imply spurious associations whereas deviation of only a few SNP markers for the observed and expected P values supports the GWAS model. Overall, both the GLM models showed large deviations for the observed and expected P values for IT and SEV in field tests of both Pakistan and the U.S. Investigation of the QQ plots revealed that the Farm CPU model containing PC3 as a correction for population structure performed better than the rest of the models and was utilized further to find SNP associations with stripe rust IT and SEV.

Genome-wide association analyses performed on the IT and SEV data measured in response to different Pst races, field tests under artificial inoculation in Pakistani and natural infection in Washington State, and corresponding BLUPs at seedling and adult-plant stage revealed 78 (IT) and 49 (SEV) significant SNP markers at minimal probability (p<0.005). Implementation of FDR adjusted value of (p<0.1) reduced the number of

significant markers to seven and four for IT and SEV, respectively. Further, all seven significant SNP markers were able to pass the Farm CPU generated p-value threshold for IT , however, only two SNP markers were able to pass the Farm CPU generated p-value threshold for SEV (Table 4). Of the seven associated SNP markers for IT, the marker-trait associations were detected only for the reaction of the genotypes to races PSTv-101 and PSTv-76 in the seedling tests. No associations were found for the IT data of the field tests under artificial inoculation in Pakistan and the IT data of the later stage in Washington based on the Farm CPU model used for GWAS. For SEV, a significant marker-trait association was found at the seedling plant stage at the Mount Vernon location and the adult plant stage at the NARC location in 2015. The highest number of associations (six) were detected for IT response to race PSTv-101 at seedling followed by three associations for SEV response under field conditions at the earlier stage at the Mount Vernon location (Fig. 5, Table 6). Six IT response SNP associations were distributed on chromosomes 1A, 2A, 2D, 5A, and 5B and the three SEV response–SNP associations were distributed on 2A, 6B, and 7B. Out of all markers one marker was significant that is present on 5A was linked with IT race response (PSTv-76). Similarly, one marker located on 7A was found associated with SEV in the field tests at the NARC location. Among all 11 significant marker-trait associations, IWB25202 showed the highest significance (–log10 (p) = 8.23) for IT to race PSTv-76 and the lowest significance (–log10 (p) = 3.86) was detected for SEV at the NARC location. The names of the SNP markers, mapped positions, favorable alleles, minor allele frequencies, and the threshold P values are provided in Table 6. IWB25202 was associated with the IT data from the tests to both races PSTv-101 and PSTv-76, though with varying significance.

Table 5. Pearson's correlation coefficients for IT and SEV of Pakistani wheat genotypes grown in Washington, US.

Environment	IT vs. IT		SEV vs. SEV		IT↓ vs. SEV→	
	Pullman	Mount vernon	Pullman	Mount vernon	Pullman	Mount vernon
Pullman	1	0.69**	1	0.77**	0.87**	0.69**
Mount Vernon	0.69**	1	0.77**	1	0.74**	0.91**

Asterisks * and ** Indicate $p<0.05$ and $p<0.001$, respectively, and ns = Not significant

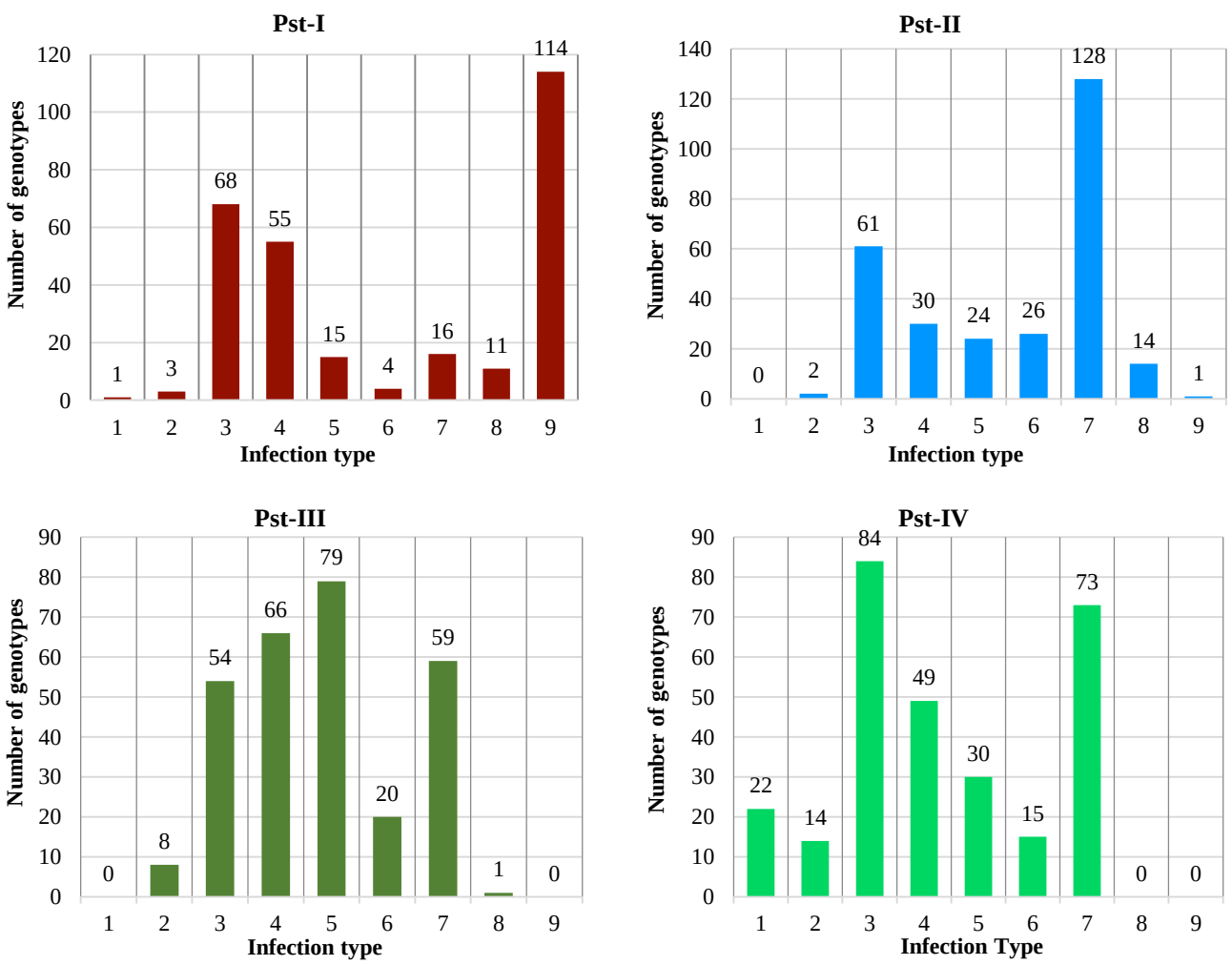


Fig. 3. Frequency distribution of genotypes evaluated at seedling stage for four *Puccinia striiformis* f. sp. *tritici* races. Genotypes were scored on a scale of 0-9 and the genotypes with a score of ≤ 3 were considered resistant.

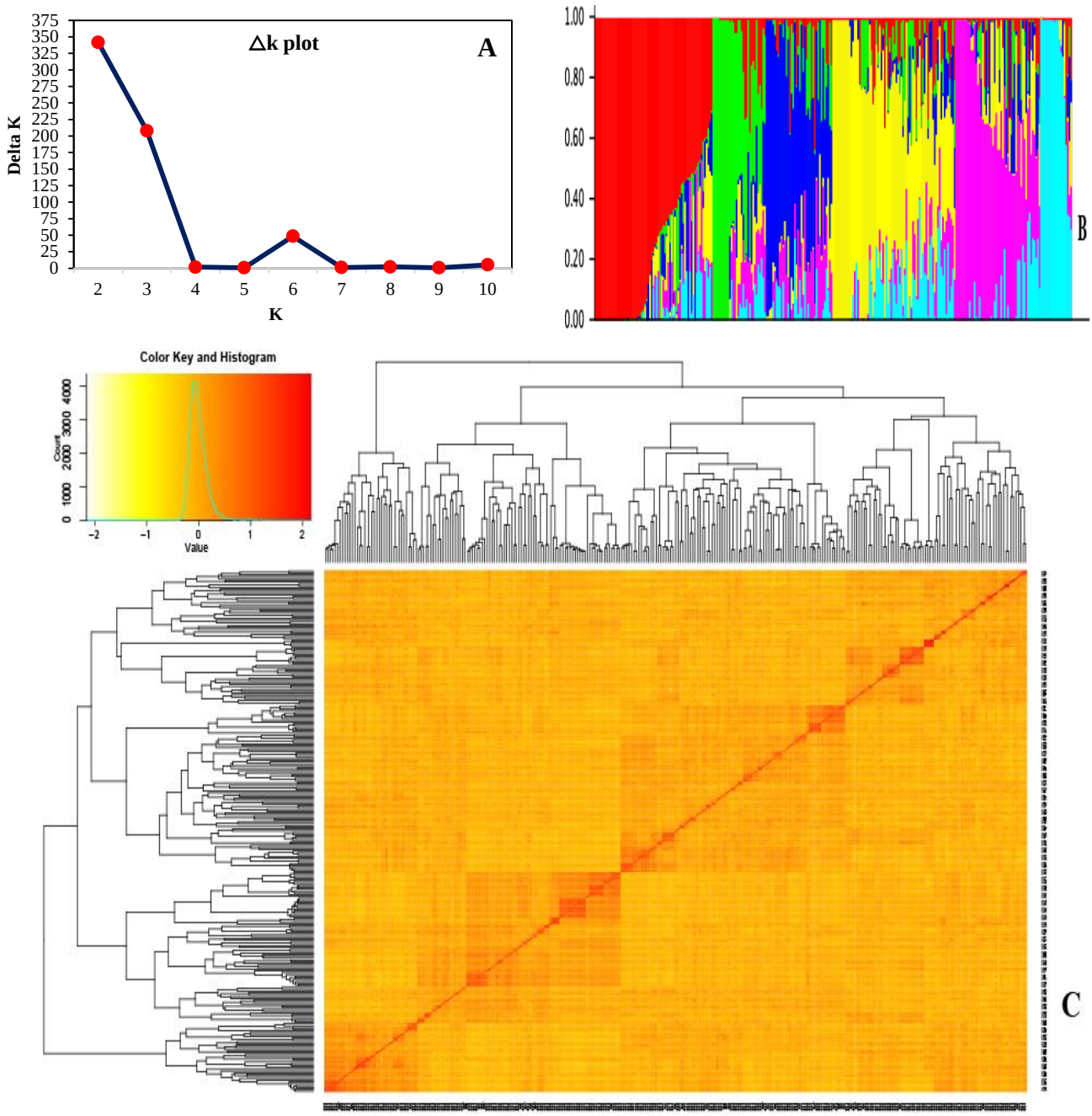


Fig. 4. Population structure and kinship analysis of Pakistani wheat germplasm. (A) Plot of delta K vs. K from 1 to 10 to estimate the best K, and the presence of peak at K=6 hints six subgroups (B) Partitioning of genotypes into six subpopulations based on STRUCTURE analysis. (C) A heat map of the Identity-by-decent kinship matrix illustrating intermediate to high levels of relatedness.

Table 6. SNPs and the beneficial allele for stripe rust resistance at seedling and adult-plant stage as identified by GWAS.

Trait	Stage	SNP marker	Chr	cM	Allele	Minor allele frequency	-log$_{10}$ (*P*-value)	FDR *P*-value	Farm CPU *P*-value	Effect
IT (Pst-I)	Seedling	IWB11756	1A	70.1	**A**/G	0.16	5.82	0.00013	0.00012	-1.04
IT (Pst-I)	Seedling	IWA5893	2A	97.5	A/**G**	0.17	4.20	0.00013	0.00012	1.12
IT (Pst-I)	Seedling	IWA2640	2A	116.2	**T**/C	0.05	4.30	0.00013	0.00012	1.16
IT (Pst-I)	Seedling	IWA1601	2D	5.9	A/**C**	0.05	5.13	0.00013	0.00012	2.08
IT (Pst-I)	Seedling	IWB25202	5A	129.8	**A**/G	0.20	5.30	0.00013	0.00012	-1.14
IT (Pst-I)	Seedling	IWB27386	5B	185.3	T/**C**	0.22	4.21	0.00013	0.00012	-0.99
IT (Pst-II)	Seedling	IWB25202	5A	129.8	**A**/G	0.20	8.23	0.00013	0.000014	-0.99
SEV (M2018)	Seedling	IWB11136	2A	9.4	**T**/G	0.40	3.90	0.00013	0.000049	3.84
SEV (M2018)	Seedling	IWA8189	6B	64.7	**A**/G	0.24	4.44	0.00013	0.000049	-5.28
SEV (M2018)	Seedling	IWA8570	7B	90.4	**A**/G	0.29	4.09	0.00013	0.000049	-4.11
SEV (N2015)	Adult-plant	IWB7063	7A	49.1	A/**G**	0.15	3.86	0.00013	0.00029	8.70

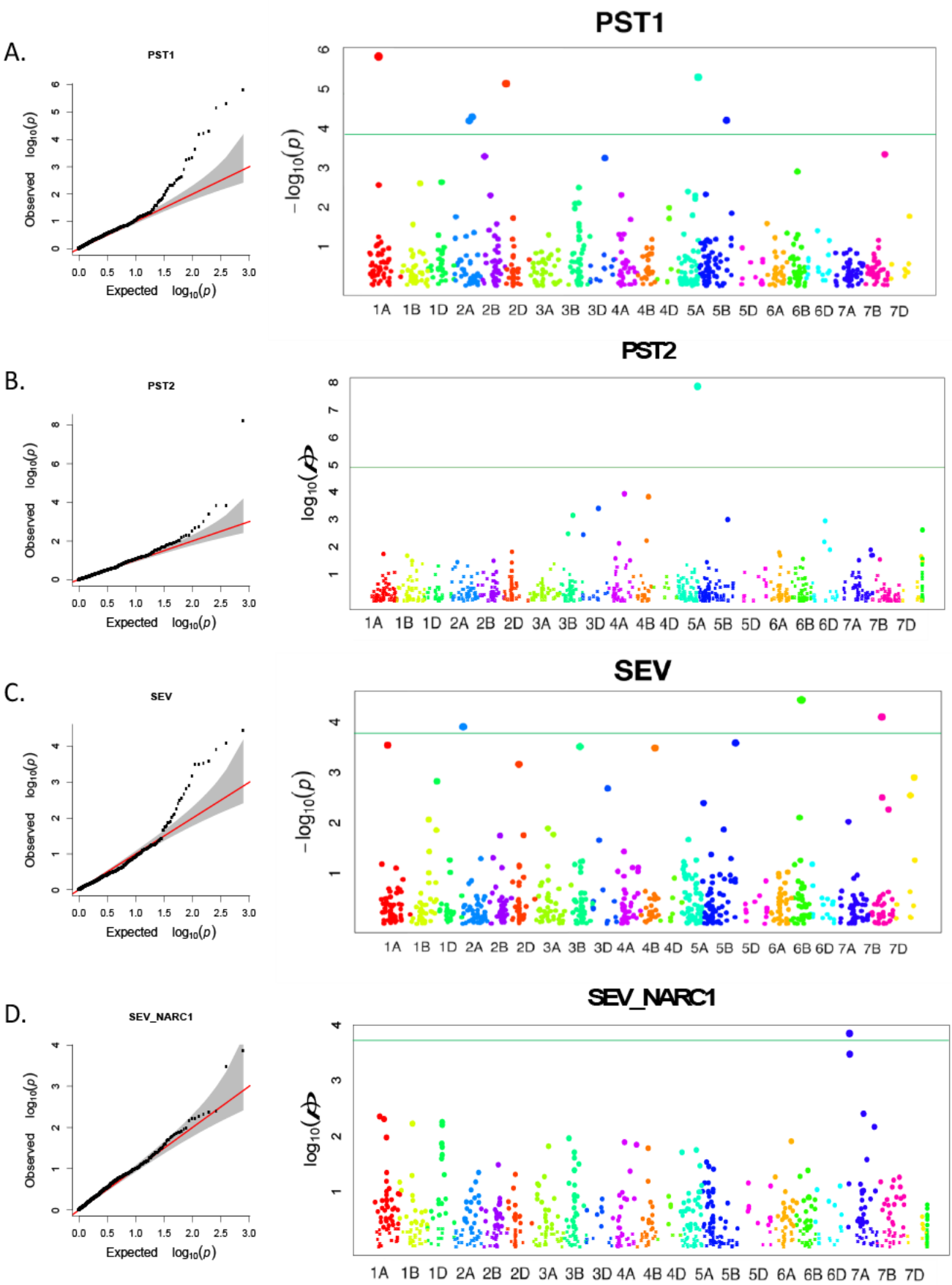
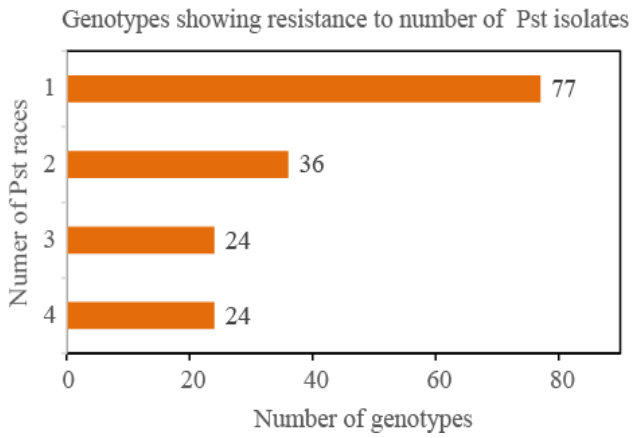


Fig. 5. GWAS-derived Quantile-Quantile plots and Manhattan plots of the FarmCPU model with PC3 as covariate for traits with significant associations. (A) Q-Q plot and Manhattan plot for resistance to Pst-I. (B) Q-Q plot and Manhattan plot for resistance to Pst-II. (C) Q-Q plot and Manhattan plot for SEV at seedling stage in the U.S. environment in 2018. (D) Q-Q plot and Manhattan plot for SEV at adult plant stage in one of Pakistani environments in 2015.

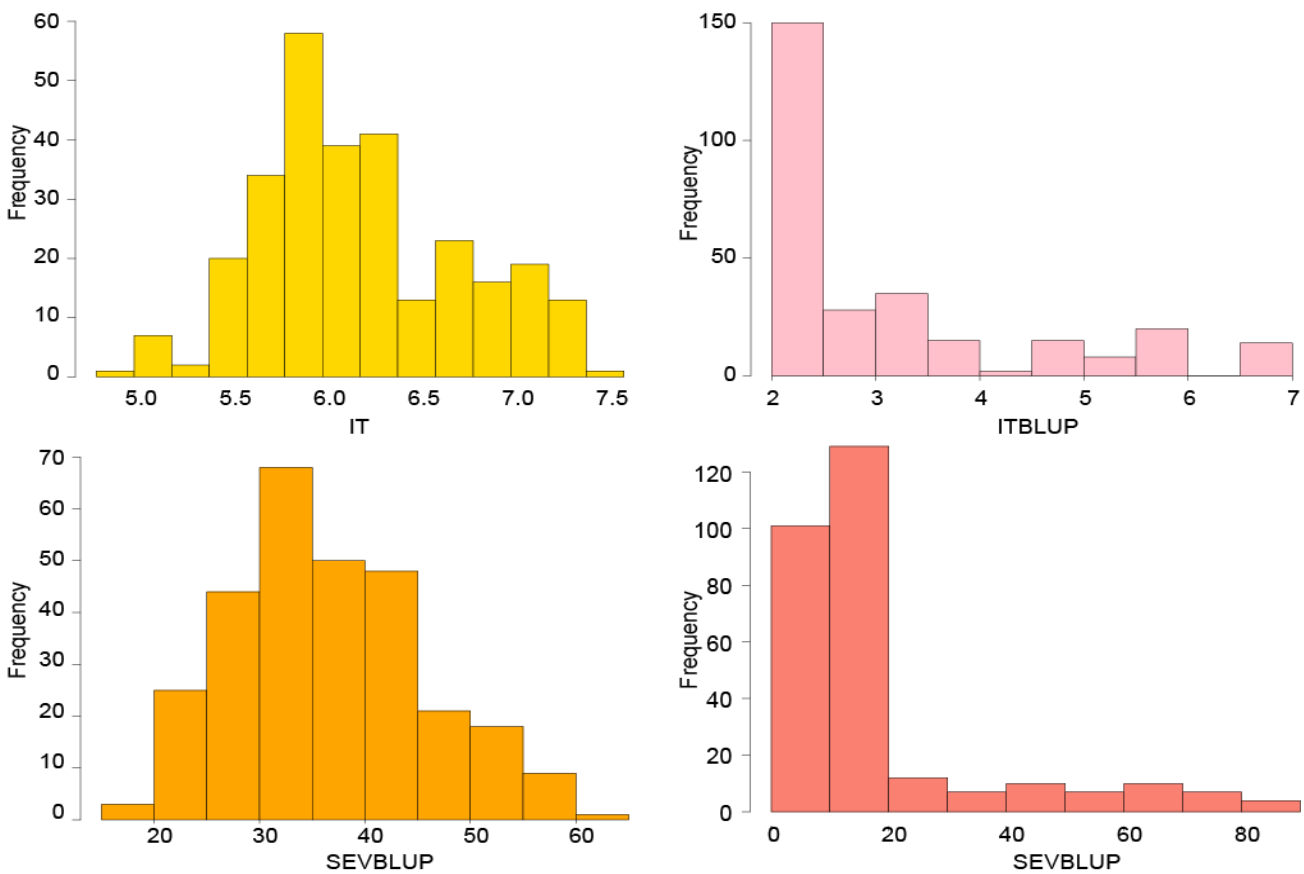


Supplementary Fig. 1. Distribution of genotypes resistant to the number of Pst races tested. All genotypes were screened for resistance to four races of Pst at seedling stage in the greenhouse.

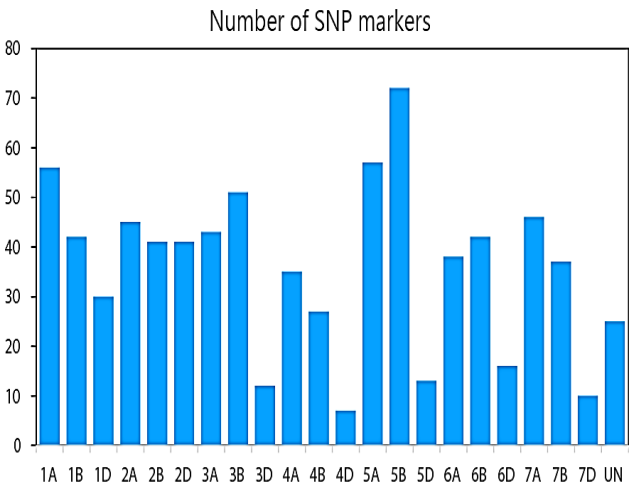
Discussion

The limited availability of effective resistance sources for stripe rust has constrained the capacity of releasing new varieties with durable resistance. The present priority for stripe resistance requires continuous determination of the gene combination status for new resistance sources that must be incorporating into present-day breeding programs for cultivar development (Khattak *et al.,* 2020). This study characterized the diverse set of 294 genotypes obtained from National Agricultural Research Centre (NARC), Islamabad for seedling and adult plant resistance against the prevailing population of 4 races of Pst in Pakistan and the Pacific Northwest of the USA. We observed a considerable amount of variation for seedling and APR in this germplasm. The avirulence/virulence of the four races is provided in Table 1. The responses of the genotypes varied greatly against the isolates, indicating that the resistance observed at the seedling stage is largely controlled by different genes that are race-specific. Out of 294 genotypes, we observed 24 genotypes being resistant to rust races (all four) of Pst at the seedling plant stage which shows that must be containing some race-specific resistance genes which can be analyzed in GWAS. Among the 97 genotypes which showed APR indicated the rust resistance is likely conversed by APR genes/QTLs. In general, genotypes that originated from both the center of diversity and origin of wheat and Pst provide the coexistence of wheat and rust pathogen in a tough natural arm race results in the addition of diverse genes in wheat for resistance to rust (Ali *et al*., 2014).

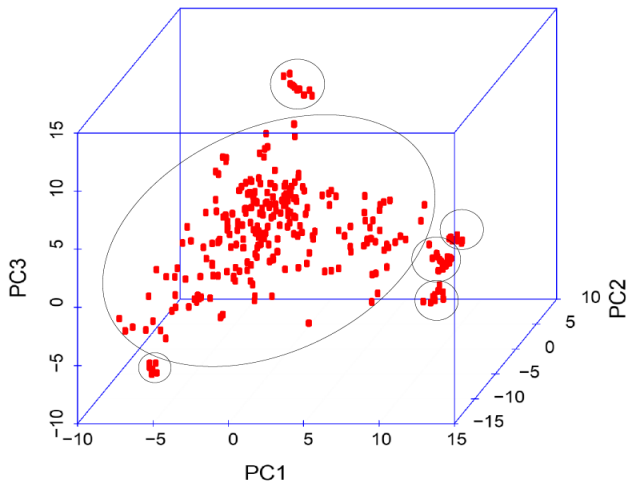
Previously PCA and SSR were used for identifying diversity at phenotypic and genetic level, thus providing a chance to plant breeders and geneticists to select desirable genotypes out of base population with unknown diversity or variability (Khattak *et al.,* 2020). But still to identify the genomic regions responsible for rust resistance GWAS analysis is the ultimately source for taking out the genes for resistance in germplasm sources and integrating them into different breeding programs will ultimately increase the resistance of newly developed cultivars which will reduce the wheat losses due to this disease.



Supplementary Fig. 2. Histograms illustrating frequency distributions IT and SEV BLUPs. (A) Distribution of IT BLUPs under artificial epidemics (CCRI and NARC). (B) Distribution of IT BLUPs under natural infection conditions (Pullman and Mt Vernon). (C) Distribution of SEV BLUPs under artificial epidemics (CCRI and NARC). (D) Distribution of SEV BLUPs under natural infection conditions (Pullman and Mt Vernon).



Supplementary Fig. 3. Chromosome-wise SNP markers distributed across 21 wheat chromosomes utilized in this study including SNPs with unknown (UN) chromosome location.



Supplementary Fig. 4. Population structure of 287 genotypes identified through the first three principal components in GAPIT program. Six subgroups were inferred.

Population structure: The analysis was performed to explain the genetic structure of the wheat germplasm used in the experiment. The structure analysis revelated the clustering of the panel of 287 wheat genotypes into six main sub-populations and revealed a significant arrangement of the pattern of subpopulations. The population structure was also performed by the principal component analysis which also groups the whole data set into six major groups. The first, second, and third PCs explained the 20, 14, and 9% of variation respectively. The first and second PCs consist of released wheat cultivars in Pakistan and lines from the National Uniform Wheat Yield Trial, respectively. Further, genotypes present in the first two clusters of PCs contain the majority of the genotypes conferring adult plant resistance. This interrelation among the population structure and resistance to stripe rust can be due to the significance of the difference between various regions of wheat growing concerning the widespread presence and prevalent pathogen population that is causing variation in genetic architecture of disease resistance.

Identification of significant QTL by GWAS: We identified 11 loci, of which 7 were for IT and 4 were for SEV. No associations were found for the IT data for the artificial inoculation field tests in Pakistan and natural field conditions in the US at the adult plant stage based on the Farm CPU model used for GWAS. This could be due to several reasons. First, the relatively small number of useful markers (787) might leave large gaps for missing resistance QTL. Second, the use of mixtures of multiple races might eliminate genes for race-specific resistance. In other words, the tested Pakistani wheat germplasm may not contain genes for effective resistance against all races used in the mixture. Third, Pakistani wheat germplasm was mostly resistant with 247 (83%) genotypes having IT 2-3 and 33 (11%) intermediate (IT 5), but only 17 (6%) susceptible (IT 8). The lack of an adequate number of susceptible genotypes in the second note-taking at the Mount Vernon location was the major reason for the failure to detect significant loci for the IT data. However, the resistance genes can be identified by crossing selected genotypes to a susceptible variety.

For SEV, a significant marker-trait association was found at the middle jointing stage at the Mount Vernon location and the adult-plant stage at the NARC location in 2015. The genetic map constructed by Maccaferri *et al*., (2015) was used to determine the prospective co-localization of the resistance loci with previously mapped genes. The genomic regions associated with 7 loci associated with resistance to IT and 4 loci associated with SEV were very closely mapped to known stripe rust resistance genes and loci reported in the previous study. This study further validated the result provided in Maccaferri *et al*., (2015) and strengthen the argument that the resistance genes among these virulent types are even present in the germplasm being used in Pakistan. The identified loci can be used in breeding operations in Pakistan to develop some of the durable cultivars.

Conclusion

Our study well supported previous studies along with authentication of precision of present association analysis. This study also identified a new locus that was highly reliable as it was identified in wheat accessions with high yield potential. By combining this with other stripe rust resistance genes, durable resistance can be easily developed. The genotypes selected can be used by breeders for developing cultivars as they were having more resistance alleles against rust. Nevertheless, the panel was able to produce a decent and clear picture of the current genetic diversity against rust resistance genes in indigenous wheat germplasms.

Acknowledgments

Sania Begum is appreciated for providing technical support at NIGAB, NARC, Islamabad. Dr. Javed Mirza is appreciated for conducting seedling stage testing at Crop Disease Research Institute Murree. Deven See, Kalvinder Gill, Kent Evans, and Muhammad Ahsan are appreciated for technical support and Guidance in performing GWAS at Washington State University.

References

Ali, S., P. Gladieux, M. Leconte, A. Gautier, A.F. Justesen, M.S. Hovmøller and C. de Vallavieille-Pope. 2014. Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen *Puccinia striiformis* f.sp. *tritici*. *PLoS Pathog.*, 10(1). https://doi.org/10.1371/journal.ppat.1003903.

Bernardo, A., P. St. Amand, H.Q. Le, Z. Su and G. Bai. 2020. Multiplex restriction amplicon sequencing: A novel next-generation sequencing-based marker platform for high-throughput genotyping. *Plant Biotech. J*., 18(1): 254-265.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btm308.

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. *Genetics*, 6(8): 2237-2253. https://doi.org/10.1534/g3.116.028407.

Chen, X.M. 2005. Review / Synthèse Epidemiology and control of stripe rust [*Puccinia striiformis* f . sp . *tritici*] on wheat. *Can. J. Bot.*, 27: 314-337. https://doi.org/10.1071/ar07045

Chen, X.M. 2013. High-temperature adult-plant resistance, key for sustainable control of stripe rust. *Amer. J. Plant Sci.*, 4: 608-627.

Chen, X.M. 2014. Integration of cultivar resistance and fungicide application for control of wheat stripe rust. *Can. J. Plant Pathol.*, 36: 311-326.

Doyle, J.J. and J.L. Doyle. 1987. *A rapid DNA isolation procedure for small quantities of fresh leaf tissue* (No. RESEARCH).

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(8): 2611-2620.

Godoy, J., S. Rynearson, X.M. Chen and M. Pumphrey. 2018. Genome-wide association mapping of loci for resistance to stripe rust in North American elite spring wheat germplasm. *Phytopathology,* 108: 234-245.

Goldstein, D.B. and M.E. Weale. 2001. Linkage disequilibrium holds the key. *Curr. Biol.*, 11(14): 576-579. https://www.fao.org/home/en/

Khattak, S.H., S. Begum, M. Aqeel, M. Fayyaz, S.A.K. Bangash, M.N. Riaz, S. Saeed, A. Ahmed and G.M. Ali. 2020. Investigating the allelic variation of loci controlling rust resistance genes in wheat (*Triticum aestivum* L.) land races by SSR marker." *Appl. Ecology and Environ. Res*., 18(6): 8091-8118.

Kollers, S., B. Rodemann, J. Ling, V. Korzun, E. Ebmeyer, O. Argillier and M.S. Röder. 2013. Genetic architecture of resistance to Septoria tritici blotch (*Mycosphaerella graminicola*) in European winter wheat. *Mol. Breed.*, 32(2): 411-423.

Kraakman, A.T.W., F. Martínez, B. Mussiraliev, F.A. Van Eeuwijk and R.E. Niks. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. *Mol. Breed.*, 17(1): 41-58.

Line, R.F. and A. Qayoum. 1992. Virulence aggressiveness, evolution, and distribution of races of *Puccinia striiformis* (the cause of stripe rust of wheat) in North America, 1968-87. U.S. Department of Agriculture Technical Bulletin No. 1788, 44 pp.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury and Z. Zhang. 2012. *GAPIT: Genome Association and Prediction Integrated Tool*, 28(18): 2397-2399.

Liu, W.Z., M. Maccaferri, X.M. Chen, M. Pumphrey, G. Laghetti, D. Pignone and R. Tuberosa. 2017. Genome-wide association mapping reveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum* ssp. *dicoccum*). *Theor. & Appl. Gen*., 130: 2249-2270.

Liu, X, H. Meng H, F. Bin, S. Edward and Z.Z. Buckler. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, 12(e1005767).

Nordborg, M. and D. Weigel. 2008. Next-generation genetics in plants. *Nature*, 456(7223): 720-723.

Peterson, R.F., A.B. Campbell and A.E. Hannah. 2011. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Can. J. Res.*, 26c(5): 496-500.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data.

Pupo, G., A.C. Vicini, D.M.H. Ascough, F. Ibba, K.E. Christensen, A.L. Thompson and V. Gouverneur. 2019. Hydrogen bonding phase-transfer catalysis with potassium fluoride: enantioselective synthesis of β-Fluoroamines. *J. Amer. Chem. Soc.*, 141(7): 2878-2883. https://doi.org/10.1021/jacs.8b12568

Rafalski, J.A. 2010. Association genetics in crop improvement. *Curr. Opin. Plant Biol.,* https://doi.org/10.1016/j.pbi.2009.12.004.

Rehman M.A., R. Saleem, S.W. Hasan, S. Inam, S.Z. Uddin, M. Saeed, S. Noor, M.N. Riaz, G.M. Ali and S.H. Khattak. 2020. Economic assessment of cereal - Legume intercropping system, a way forward for improving productivity and sustaining soil health. *IJBPAS*, 9(5): 1078-1089.

Sela, H., S. Ezrati, P. Ben-Yehuda, J. Manisterski, E. Akhunov, J. Dvorak and A. Korol. 2014. Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) population in Israel. *Theor. & Appl. Gen.*, 127(11): 2453-2463.

Tadesse, W., F.C. Ogbonnaya, A. Jighly, M. Sanchez-Garcia, Q. Sohail, S. Rajaram and M. Baum. 2015. Genome-wide association mapping of yield and grain quality traits in winter wheat genotypes. *PLoS ONE*, 10(10): 1-13.

Vanraden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.*, 91(11): 4414-4423.

Wan, A.M., X.M. Chen and J. Yuen. 2016. Races of *Puccinia striiformis* f. sp. *tritici* in the United States in 2011 and 2012 and comparison with races in 2010. *Plant Dis.,* 100: 966-975.

Wang, M.N. and X.M. Chen. 2017. Stripe rust resistance. Pages 353-558 In: (Eds.): Chen, X.M. & Z.S. Kang. Stripe Rust. Springer, Dordrecht.

Waqar, A., S.H. Khattak, S. Begum, T. Rehman, A. Shehzad, W. Ajmal and G.M. Ali. 2018. Stripe rust: A review of the disease, YR genes and its molecular markers. *Sarhad Journal of Agriculture*, 34(1):

William, M., R.P. Singh, J. Huerta-Espino, S.O. Islas and D. Hoisington. 2003. Molecular marker mapping of leaf rust resistance gene lr46 and its association with stripe rust resistance gene yr29 in wheat. *Phytopat.*, 93(2): 153-159.

Zadoks, J.C., T.T. Chang and C.F. Konzak. 1974. A decimal code for the growth stages of cereals. *Weed Research*, 14(6): 415-421.

Zegeye, H., A. Rasheed F. Makdis, A. Badebo and F.C. Ogbonnaya. 2014. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS ONE*, 9(8): 1-18.

(Received for publication 8 October 2021)