

Association mapping of Iranian wheat genotypes to detect QTLs conferring resistance to stem rust at seedling stage

Ali Saremirad

Karaj Islamic Azad University

Mohammad Bihamta (✉ mrghanad@ut.ac.ir)

University of Tehran

Ali Malhipour

Seed and Plant Improvement Research Institute

Khodadad Mostafavi

Karaj Islamic Azad University

Hadi Alipour

Urmia University

Research article

Keywords: Wheat, Association mapping, Puccinia graminis f. sp. tritici, Sr genes, SNP markers

Posted Date: May 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-25893/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Association mapping of Iranian wheat genotypes to detect QTLs conferring resistance to stem rust at seedling stage

Abstract

Background: Stem rust caused by *Puccinia graminis* f. sp. *tritici* (*Pgt*) is an important disease of wheat in the world. *Pgt* pathogen is constantly evolving and creating more virulent races that break down stem rust (*Sr*) resistance genes. As a result, many of *Sr* genes have become ineffective against new *Pgt* races. Exploring new sources of resistance to detect new *Sr* genes/QTLs is very important in order to introducing them into wheat breeding programs and developing resistant wheat cultivars. The objective of the present study was to evaluate 297 Iranian wheat genotypes for resistance to stem rust at seedling stage and to detect *Sr* resistance genes/QTLs through association mapping (AM).

Results: A set of 297 Iran bread wheat cultivars and landraces were evaluated for infection type and latent period in four race of *Pgt*. Genotypic data of 282 genotypes were available, so AM was performed based on 282 genotypes. The results of population structure analysis showed that 277 genotypes clearly were distinguished in the three subpopulations and the other five genotypes were classified in the mixed group. The mean linkage disequilibrium decreased with increasing genetic distance. The markers did not have a uniform distribution on the genomes, so the share of each of the A, B and D genomes in commercial cultivars and landraces was approximately 37, 46 and 17%, respectively. Collectively, 69 QTLs for infection type and 62 QTLs for latent period of studied *Pgt* races were identified in the original dataset ($P \leq 0.001$). In the imputed SNPs dataset, the number of QTLs for infection type increased to 504 QTL and for latent period increased to 454 QTLs ($P \leq 0.001$).

Conclusion: Based on the results of this study, it can be concluded that the Iranian wheat genotypes are valuable source resistance to stem rust. By incorporating these genotypes into wheat breeding programs and optimizing effective resistance genes, an important step can be

taken to prevent the threat of and the disease to ensure food security. This study provides additional useful information for selection of resistant genotypes against the disease by improving marker-assisted selection efficiency.

Keywords: Wheat, Association mapping, *Puccinia graminis* f. sp. *tritici*, *Sr* genes, SNP markers.

Background

In cereal classification, wheat with more than 221 million hectares of cultivation area and more than 751 million tons of grain production annually, is recognized as the third most important cereal worldwide [1]. In Iran, approximately 7.65 million hectares out of about 11 million hectares of field crops in 2016-2017 cropping year has been devoted to cereal cultivation and of the total grain harvest, 70.22% to wheat, 20.84% to barley, 0.76% to rice paddy and 1.88% to grain corn was belonged [2]. These statistics show the great importance of grains, especially wheat, in the country. It is predicted that a 50% increase in cereal yield will be needed to meet the nutritional needs of the human population by 2050 [3]. Biotic and abiotic stresses challenge grain production and yield increase in wheat. Diseases and pests are among the most important biotic stresses that reduce approximately 20% of the world wheat yield annually [4]. Rusts are one of the most dangerous cereal diseases that have coexisted and evolved during the use of cereals [5]. Black or stem rust of wheat caused by *Puccinia graminis* Pers. f. sp. *tritici* (*Pgt*) is considered a very serious threat in major wheat growing areas around the world. This disease, which usually develops under humid conditions and high temperatures (15 to 35 °C), is the most damaging rust among wheat rusts and the extent of damage may result in the complete destruction of large-scale wheat fields [6, 7].

Application of fungicides and use of resistant wheat cultivars are two important strategies to combat *Pgt*. However, use of fungicides may not always be feasible due to the high cost, risk of fungal resistance to the fungicides, and the possibility of environmental contamination. Use of resistant wheat cultivars is an effective and economical way of controlling the disease. In this regard, it is important to have sufficient knowledge of population genetics of the causal agent of the disease and identification of resistance genes effective in wheat genotypes to be selected.

Based on the inheritance in plants, resistance to disease are monogenic and polygenic. In the monogenic state, a gene is responsible for the expression of resistance, and the type of resistance, dominance and codominance, and discrete variation can be obtained in differentiating generations and can be identified using the Mendelian rules and the monosomic technique of the genes. In polygenic state, resistance is controlled by several genes, so monosomic techniques and Mendel's rules cannot be used to identify resistance genes. Linkage mapping to detect quantitative traits loci (QTLs) is one of the methods used for the genetic study of traits controlled by multiple genes. The major limitations of this method are limited number of crossing overs, resulting in low resolution of the genetic map (10-20 cM) and high cost of population replication to reach sufficient crossing overs. In addition, the created populations are only useful for limited traits and studies [8, 9]. An alternative method that has the advantage of using natural populations is association mapping. The purpose of association mapping is to identify continuous markers linked to one or more quantitative traits. Association mapping is based on the assumption that the observed phenotypic variation is related to genetic variation. In this method, a large and diverse set of individuals from a population is randomly collected and mapping is done based on linkage disequilibrium. Association mapping has been used for both whole-genome scanning and candidate gene analysis [10-14]. Association mapping is much more accurate than linkage

mapping due to its many recombination's [15]. In addition, it is more compatible with genetically diverse germplasm and allows mapping of several traits at the same time. Therefore, for each trait desired, there is no need to create two-parental populations, which in turn incurs additional costs for genotypic and phenotypic evaluation.

The objectives of this study were to carry out a genome wide search in Iranian wheat genotypes for resistance loci to *Pgt* races at the seedling stage, and the identification of genomic regions suitable for marker-assisted selection and further genetic dissection.

Results

Phenotypic data analysis

Combined analysis of variance was performed for data of infection types and latent period as Bartlett test [16] confirmed homogeneity of variance of experimental errors for data. According to the results, significant variation ($P \leq 0.01$) was observed among the wheat genotypes and *Pgt* races in terms of infection types and latent period. The effect of genotype \times race was also significant ($P \leq 0.01$) for both characteristics of infection types and latent period which indicates that for the different *Pgt* races, the response of the genotypes varied in terms of infection types and latent period (Table 1). Simple variance analysis of infection types and latent period for each race of *Pgt* was performed to detect variation in reaction of the genotypes to four *Pgt* individual races (Table 2). Based on these results, significant variation ($P \leq 0.01$) was observed among genotypes for infection types and latent period of every race tested.

Genetic parameters including variance components, coefficients of genetic and phenotypic variations, heritability, and genetic efficiency for infection types and latent period in wheat genotypes against four *Pgt* individual races were calculated (Table 3). The results showed high genetic and phenotypic variances for infection types and latent period of wheat

genotypes to all races Except TKTTF infection type and latent period, the rest had high genetic and phenotypic variance. While the highest genetic and phenotypic coefficients of variation were obtained for infection types and latent period of the race TTKTK, the lowest values of these coefficients were detected for TKTTF. The heritability and genetic efficiency of the infection types varied from 84.02 to 97.80% and 25.33 to 38.04%, respectively and for latent period varied from 67.68 to 95.34% and 24.85 to 33.55%, respectively.

Table 1. Combined analysis of variance of data of infection types and latent period from the reaction of 297 wheat genotypes to four *Pgt* races

Source of variation	df	Mean squares		Explained variance (%)	
		Infection type	Latent period	Infection type	Latent period
Race	3	148.03**	178.88**	3.75	7.85
Block (race)	4	18.99	3.67	0.64	0.21
Genotype	296	22.59**	11.49**	56.55	49.79
Genotype×race interaction	888	4.58**	2.53**	34.40	32.98
Error	1184	0.46	0.52	4.64	9.15
Coefficient of variation (%)		5.98	9.36		

** : Significant at 1% probability level.

Table 2. Simple analysis of variances of data of infection types and latent period from the reaction 297 wheat genotypes to four *Pgt* individual races

Source of variation	df	Mean squares							
		TTTTF		PTRTF		TTKTK		TKTTF	
		IT	LP	IT	LP	IT	LP	IT	LP
Block	1	4.37**	0.43 ^{ns}	0.13 ^{ns}	0.006 ^{ns}	71.44**	13.63**	0.02 ^{ns}	0.60 ^{ns}
Genotype	296	8.50**	4.30**	9.85**	4.20**	10.64**	7.11**	7.33**	3.47**
Error	296	0.24	0.20	0.10	0.10	0.92	1.37	0.57	0.42
CV (%)		4.20	6.02	2.91	4.36	9.00	13.74	6.47	8.50

IT: Infection type, LP: Latent period

ns and **: Non-significant and significant at 1% probability level, respectively.

Table 3. Amounts of genetic parameters of data of infection types and latent period collected from the reaction of 297 wheat genotypes to four *Pgt* individual races

genetic parameters	TTTTF		PTRTF		TTKTK		TKTTF	
	IT	LP	IT	LP	IT	LP	IT	LP
Genetic variance	4.131	2.050	4.872	2.050	4.862	2.870	3.380	1.525
Genetic coefficient of variation (%)	17.33	19.087	19.47	19.632	20.66	19.826	15.66	16.034
Phenotypic variance	4.375	2.250	4.981	2.150	5.786	4.240	3.958	1.945
Phenotypic coefficient of variation (%)	17.84	19.987	19.68	20.105	22.54	24.097	16.95	18.107
Environmental variance	0.244	0.200	0.109	0.100	0.924	1.370	0.578	0.420
Heritability (%)	94.43	91.11	97.80	95.34	84.02	67.68	85.40	78.40
Genetic efficiency (%)	29.48	31.87	33.69	33.55	33.14	28.54	25.33	24.85

IT: Infection type, LP: Latent period

The frequency of genotypes was calculated based on the 0- 4 infection type scale [17, 18] for each race and the results are shown in Table 4. Genotypes with seedling response between 0 and 2 (0- 0; - ; - ; 1- 1- 1+ - 2- 2C- 2 and 2⁺) as resistance reaction and genotypes with infection type between 3 and 4 (3- 3⁺ and 4) were considered as susceptible reaction [19]. The results showed that in TTTTF, 16 cultivars (17.98%) and 34 landraces (16.34%), in PTRTF, 18 cultivars (20.22%) and 41 landraces (19.71%), in TTKTK, 34 cultivars (38.20%) and 65 landraces (31.25%) and in TKTTF, 16 cultivars (17.98) and 36 landraces (17.31%) had resistance reaction.

Cluster analysis of four *Pgt* races based on infection types of the wheat genotypes performed using the Ward's method (Figure 1). The dendrogram obtained from the analysis classified the races into three groups with the cut line in the 5 coefficient. The TTTTF and TKTTF races had the least frequency of resistance reaction (Table 4). The cause of slight difference in the frequency of the TKTTF resistance reaction to the TTTTF race can be attributed to the presence of the *Sr11* gene, while the resistance of this gene is broken by the TTTTF race. The other two races (PTRTF and TTKTK) showed independent virulence patterns and each was housed in a separate group.

Table 4. Classification of 297 wheat genotypes based on reaction to four individual *Pgt* races

Infection type	Cultivar/ landrace	TTTTF		PTRTF		TTKTK		TKTTF	
		Number	%	Number	%	Number	%	Number	%
Resistance reactions:									
0, 0; and ;	Cultivar	-	-	-	-	6	6.74	1	1.12
	landrace	2	0.96	3	1.44	5	2.40	1	0.48
;1, 1 and 1 ⁺	Cultivar	4	4.49	6	6.74	11	12.36	6	6.74
	landrace	12	5.76	13	6.25	17	8.17	10	4.81
2-, 2C, 2 and 2 ⁺	Cultivar	12	13.48	12	13.48	17	19.10	9	10.11
	landrace	20	9.61	25	12.02	43	20.67	25	12.02
Total	Cultivar	16	17.98	18	20.22	34	38.20	16	17.98
	landrace	34	16.34	41	19.71	65	31.25	36	17.31
Susceptible reactions:									
3 and 3 ⁺	Cultivar	48	53.93	48	53.93	46	51.68	30	33.71
	landrace	70	33.65	92	44.23	118	56.73	64	30.77
4	Cultivar	25	28.09	23	25.84	9	10.11	43	48.41
	landrace	104	50.00	75	36.06	25	12.02	108	51.92
Total	Cultivar	73	82.02	71	79.77	55	61.80	73	82.02
	landrace	174	83.65	167	80.29	143	68.75	172	82.69

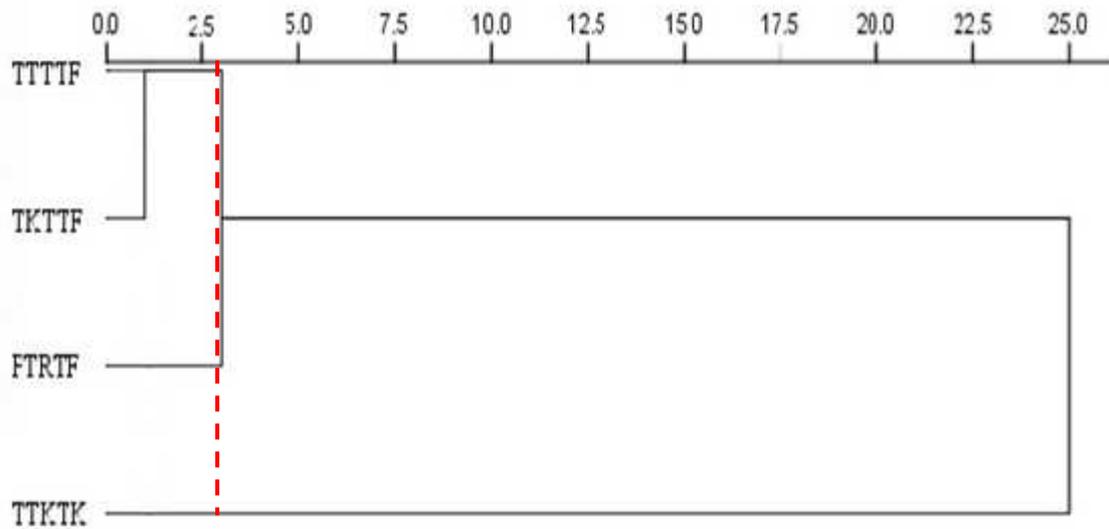


Figure 1. Dendrogram of cluster analysis of four *Pgt* races based on infection types and latent periods of 297 wheat genotypes using the Ward's method

Population structure and kinship analysis

In order to avoid of mendacious positive results, the population of the wheat genotypes was studied in terms of structure and relationship, extracting its K and ΔK statistics and plotting the 2D graph (Figure 2). The graph of ΔK at $K=3$ represents the highest value in perfectly specified fracture, dividing, the present population into three subpopulations (Figure 3). The value of membership threshold was set to 0.5.

Principal component analysis was performed on the matrix derived from the original SNPs and imputed dataset to further evaluation of population structure and investigation of genetic relationships among wheat genotypes. Based on genotypic data, the 2D plot of original (Figure 4A) and imputed SNPs (Figure 4B) showed almost distinct grouping into three groups. Genotyping variance was explained by the first two main principal components that were 12.70 and 5.95, respectively for the original SNPs and 17.20 and 6.10, respectively for imputed dataset.

Out of total of 282 genotypes, 79 genotypes (28.02%) were in the first group, 123 genotypes (43.62%) were in second group and 75 genotypes (26.59%) were in third group. The first

group contain of two cultivars and 77 landraces, the second group contain 12 cultivars and 111 landraces and the third group contain 73 cultivars and three landraces (Figure 5A). When used the imputed SNPs, the germplasms under study was also divided to three main group such as the original dataset that the first group consisted of 79 genotypes contain of two cultivars and 77 landraces and the second group contain 123 genotypes consisting of 12 cultivars and 111 landraces and the third group contain of 75 genotypes consisting of three cultivars and 72 landraces (Figure 5B).

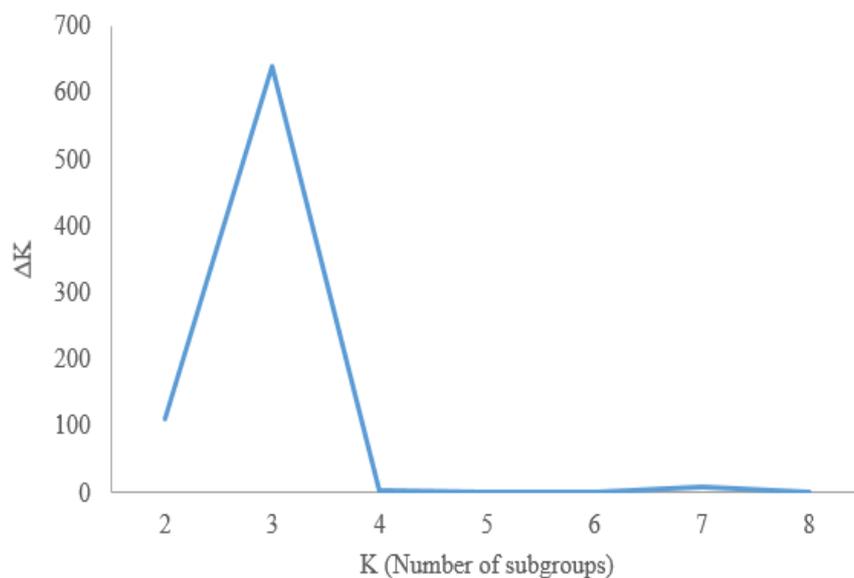


Figure 2. Determination of subpopulations number in wheat genotypes based on ΔK values

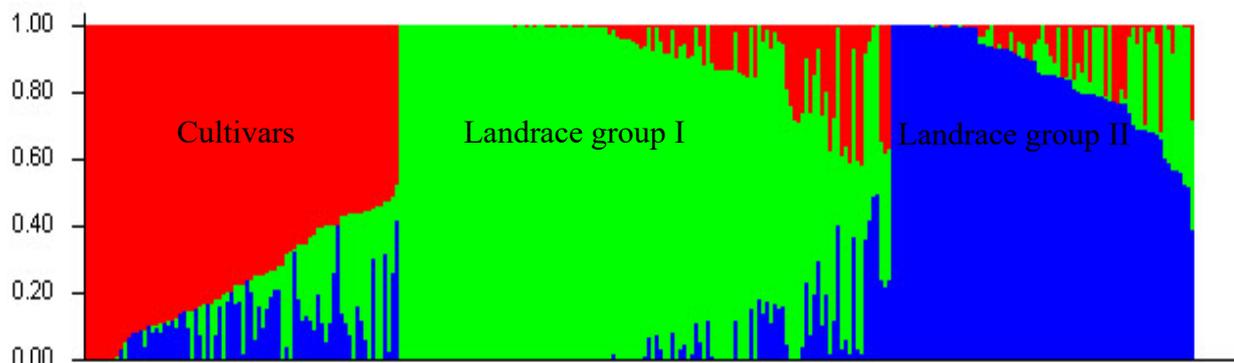


Figure 3. A structure plot of the 282 wheat genotypes and landraces determined by $K = 3$.

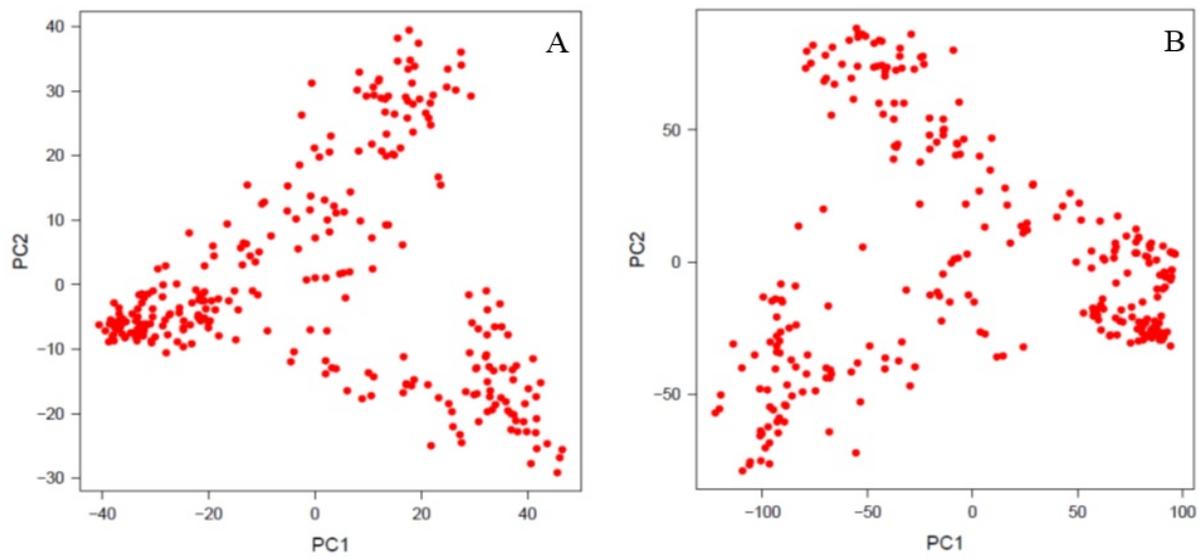
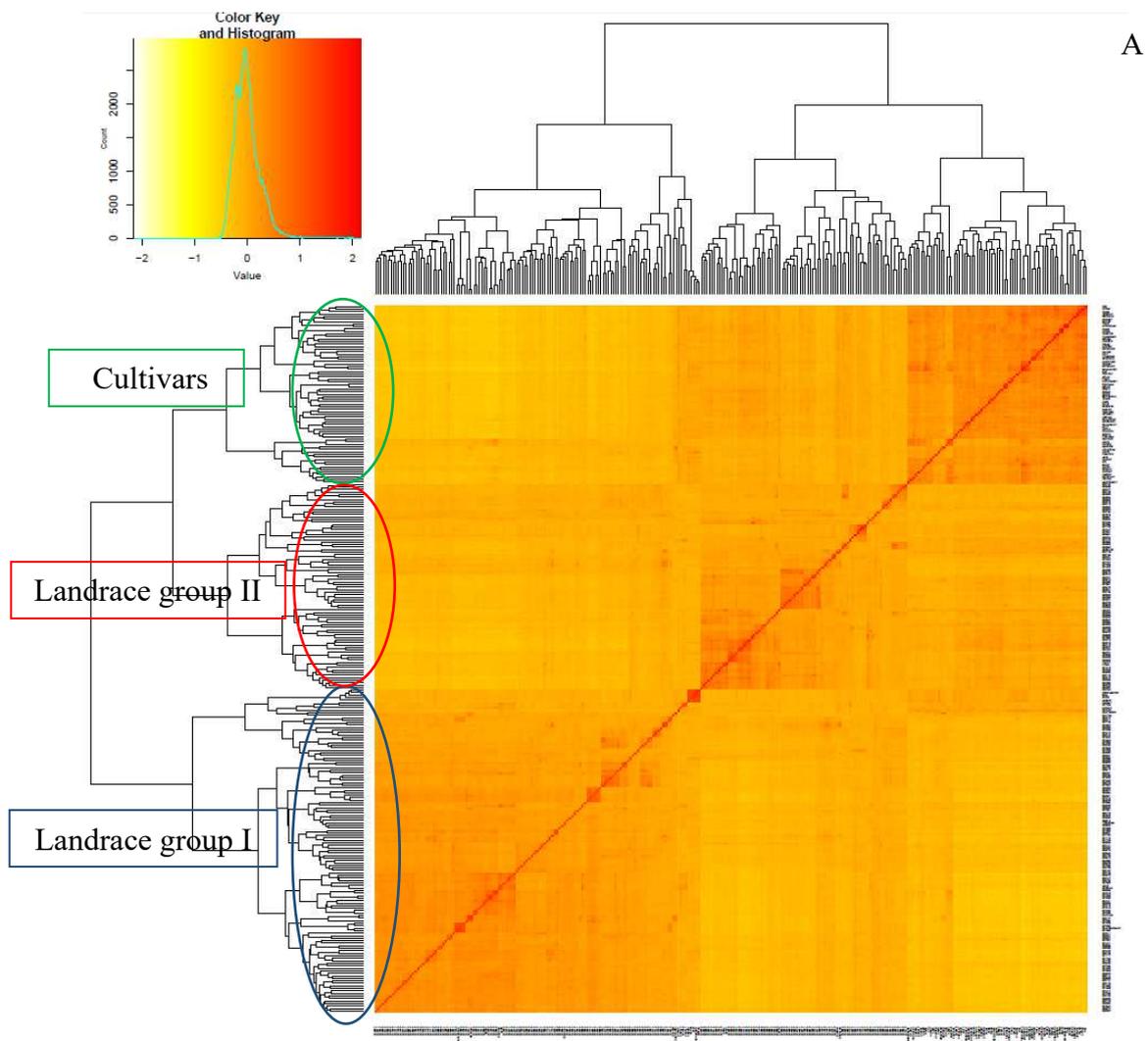


Figure 4. Principal component analysis of the wheat genotypes using original (A) and imputed SNPs (B)



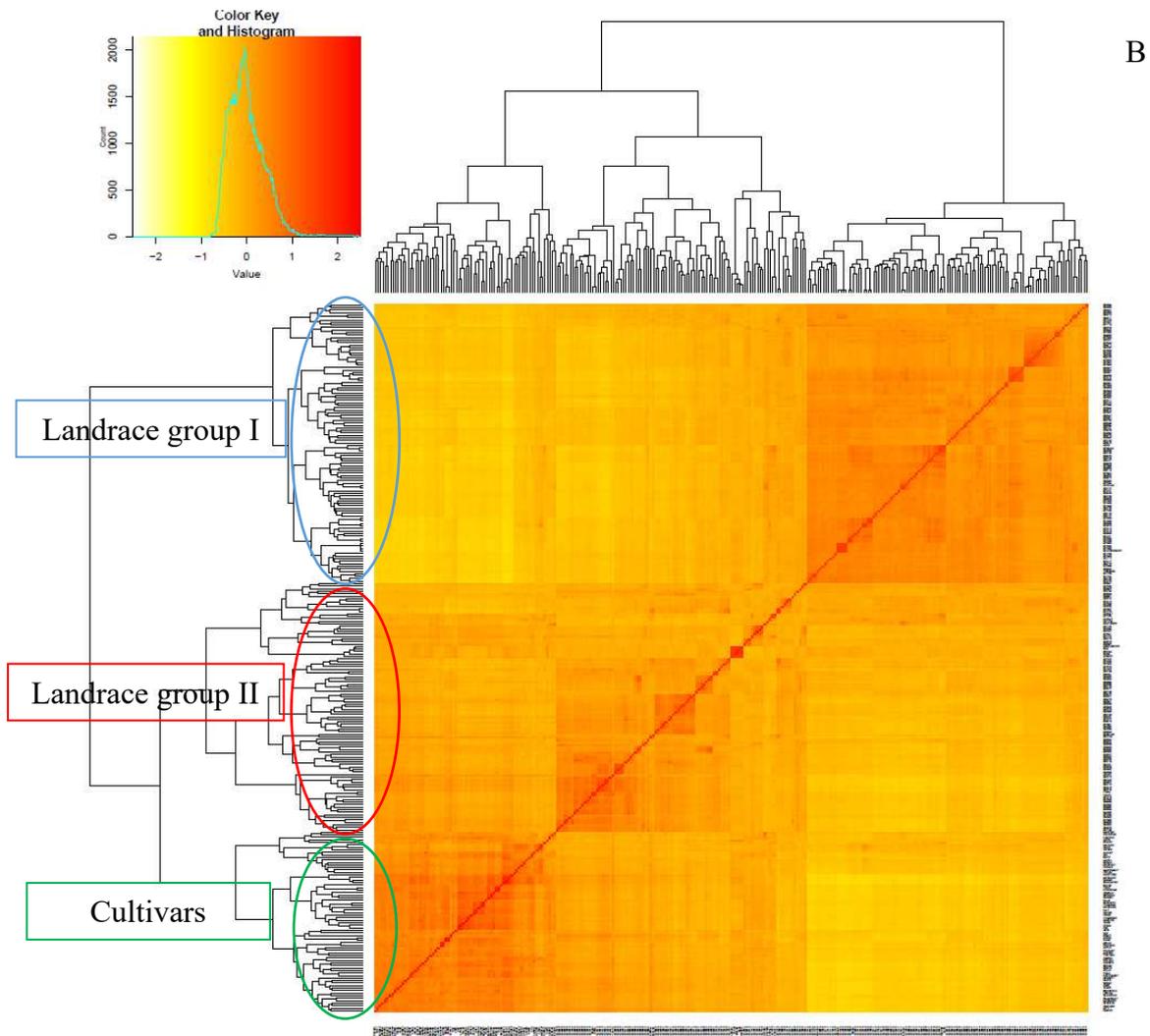


Figure 5. Cluster analysis using Kinship matrix of original (A) and imputed dataset (B) for the wheat genotypes

Linkage disequilibrium analysis

Linkage disequilibrium (LD) was examined for the wheat cultivars and the landraces by 404825 and 40824 pairwise SNPs, respectively using original dataset, and by 2152925 SNPs for both wheat cultivars and landraces using imputed dataset. The SNPs were distributed across all 21 chromosomes. The range of LD between the pairwise markers was estimated with square of allelic frequencies (r^2) between the markers. Linkage disequilibrium analysis using original dataset showed that average genetic distance for the wheat cultivars and landraces was 221.7440 and 221.7461 cM and the mean of squared allele-frequency correlation was 0.130 and 0.096, respectively (Table 5). Using imputed dataset, the average

genetic distance was equal for both wheat cultivars and landraces (52.82 cM), but their mean of squared allele-frequency correlation was 0.1924 and 0.1623, respectively (Table 6). The markers detected on different genomes were not evenly distributed. In original dataset, the number of markers located on genomes A and B in commercial cultivars were equal to the number of markers located on genomes A and B in landraces; but in D genome, landraces had a lower marker than commercial cultivars. In imputed dataset, there was no difference between the number of markers placed on each of the three genomes (A, B and D) in the commercial cultivars with the number of markers located in the three genomes (A, B and D) of landraces. The highest number of marker in original and imputed dataset was on B genome and the lowest number of marker was on D genome. The proportion of each A, B and D genomes from total pairwise original SNP markers in commercial cultivars and landraces was estimated almost 37, 46 and 17% respectively and in the imputed SNP markers approximately 36, 50 and 13%, respectively. In commercial cultivars out of 404825 pairwise SNP markers, 86415 pairwise markers (21.35%) and in landraces of 404824 pairwise SNP markers, 107976 pairwise SNP markers (26.67%) showed significant linkage at 1% probability level in original dataset. In the imputed dataset, out of 2152925 pairwise SNP markers in commercial cultivars, 842785 pairwise SNP markers (39.15%) and in landraces out of 2152925 pairwise SNP markers, 968270 pairwise SNP markers (44.98%) had significant linkage at 1% probability level.

Table 5. A summary of observed linkage disequilibrium among markers pairwise and the number of significant markers pairwise per chromosome and genome using original dataset

Chromosome	Cultivars				Landraces			
	NSP	Distance (cM)	R ²	SSP	NSP	Distance (cM)	R ²	SSP
1A	20325	6.5208	0.1095	4019	20325	6.5208	0.0731	4762
1B	26450	4.5786	0.1373	5867	26450	4.5786	0.0955	8673
1D	13350	11.9226	0.1803	3453	13350	11.9226	0.1183	4137
2A	23950	4.8296	0.1315	5427	23950	4.8296	0.1276	8259
2B	33350	4.2026	0.1288	8539	33350	4.2026	0.0966	9575
2D	17050	6.5433	0.2647	4553	17050	6.5433	0.1999	5078
3A	21500	9.5170	0.0970	3613	21500	9.5170	0.0706	4114
3B	33050	4.5077	0.1281	7928	33050	4.5077	0.0932	9166
3D	7550	22.2324	0.0933	701	7550	22.2324	0.0927	1547

4A	17450	8.0520	0.1312	3752	17450	8.0520	0.1106	4245
4B	10600	9.9421	0.0995	1847	10600	9.9421	0.0527	1386
4D	3700	25.9186	0.1359	588	3700	25.9186	0.1116	1010
5A	18250	8.0378	0.1085	3289	18250	8.0378	0.0816	4711
5B	28850	6.5358	0.1314	7721	28850	6.5358	0.0726	6691
5D	8350	27.4992	0.0936	805	8350	27.4992	0.0696	1394
6A	17850	7.3937	0.1057	3279	17850	7.3937	0.1210	6715
6B	27850	4.3847	0.1409	6565	27850	4.3847	0.0753	6593
6D	9250	18.1532	0.1055	1397	9250	18.1532	0.0802	2059
7A	28550	6.1499	0.1520	5580	28550	6.1499	0.1123	8528
7B	26800	4.7642	0.1118	5429	26800	4.7642	0.0837	6846
7D	10750	20.0585	0.1618	2063	10749	20.0606	0.0932	2487
A genome	147875	50.5007	0.1194	28959	147875	50.5007	0.0995	41334
B genome	186950	38.9156	0.1254	43896	186950	38.9156	0.0814	48930
D genome	70000	132.3277	0.1479	13560	69999	132.3298	0.1094	17712
Total	404825	221.7440	0.1309	86415	404824	221.7461	0.0968	107976

NSP: Number of SNP pairwise, SSP: Significant SNP pairwise (P<0.01)

Table 6. A summary of observed linkage disequilibrium among markers pairwise and the number of significant markers pairwise per chromosome and genome using imputed dataset

Chromosome	Cultivars				Landraces			
	NSP	Distance (cM)	R ²	SSP	NSP	Distance (cM)	R ²	SSP
1A	110675	1.3446	0.1392	33200	110675	1.3446	0.1079	41497
1B	149000	0.9442	0.1923	59280	149000	0.9442	0.1497	74142
1D	47950	3.5143	0.3028	20528	47950	3.5143	0.2257	24767
2A	136250	0.8628	0.2672	64307	136250	0.8628	0.2766	76493
2B	185800	0.7705	0.1841	77464	185800	0.7705	0.1554	86029
2D	67250	1.6415	0.2201	21080	67250	1.6415	0.1777	26828
3A	95500	2.3004	0.1487	32745	95500	2.3004	0.1212	33346
3B	199900	0.7787	0.2307	92545	199900	0.7787	0.2029	103406
3D	35850	4.6331	0.1521	7560	35850	4.6331	0.1547	12763
4A	129850	1.3855	0.3447	66704	129850	1.3855	0.3293	69407
4B	59800	2.2142	0.1378	18268	59800	2.2142	0.0968	16220
4D	13200	9.3335	0.1573	2876	13200	9.3335	0.1239	4212
5A	71350	2.0200	0.1620	24756	71350	2.0200	0.1344	28504
5B	151500	1.3002	0.1888	66026	151500	1.3002	0.1377	68944
5D	31450	6.9574	0.1470	7085	31450	6.9574	0.1286	9923
6A	98550	1.3027	0.1721	37415	98550	1.3027	0.1688	48695
6B	188750	0.6656	0.1834	75156	188750	0.6656	0.1275	82520
6D	37800	4.1908	0.1357	9535	37800	4.1908	0.1358	14802
7A	147900	1.1793	0.2230	58549	147900	1.1793	0.1951	71306
7B	148700	0.9965	0.1464	53428	148700	0.9965	0.1154	57384
7D	45900	4.4855	0.2044	14278	45900	4.4855	0.1436	17082
A genome	790075	10.39523	0.2081	317676	790075	10.39523	0.1905	369248
B genome	1083450	7.669932	0.1805	442167	1083450	7.669932	0.1408	488645
D genome	279400	34.75617	0.1885	82942	279400	34.75617	0.1557	110377
Total	2152925	52.82133	0.1924	842785	2152925	52.82133	0.1623	968270

NSP: Number of SNP Pairwise, SSP: Significant SNP Pairwise (P<0.01)

Marker-trait association

Association mapping was used to determine the association between the phenotypic data of infection type and latent period from the evaluation of seedling stage of wheat genotypes to *Pgt* and genotypic data. Table 7 shows the results of the marker-trait association for the two

traits of infection type and latent period of each studied *Pgt* race. Totally 69 marker-trait association were identified for infection type and 62 marker-trait association for latent period of race under study in the original dataset with a probability level of 0.1% ($P \leq 0.001$). In the imputed dataset, the number of marker-trait association for infection type increased to 504 and for the latent period increased to 454 ($P \leq 0.001$). The infection types of all races in the original and imputed dataset had the highest number of marker-trait association on the B genomes of chromosomes. The TTKTK infection type had the same number of marker-trait association on all three genomes in the original dataset. In the TTTTF, the lowest marker-trait association in original and imputed SNPs in genome A and in the other races in genome D were observed. Regarding the TTTTF, PTRTF, and TKTF latent periods, it was found that the highest number of marker-trait association of original SNPs was located on genome B and for TTKTK on genome A. These results were a little different in the imputed SNPs, in PTRTF the highest number of marker-trait association in genome A and in TTKTK in genome B were observed. Both the TTTTF and TKTF had the highest number of marker-trait association located on the B genome in imputed SNPs, as the results of the original SNPs. Among the infection types of races in the original and imputed SNPs, TTTTF had 29 and 208 marker-trait association respectively, and among the latent periods of races in the original and imputed SNPs, the TTKTK breed had 19 and 160 marker-trait association in the 0.1% probability was the highest among the other races.

Table 7. A summary of marker-trait associations for infection type and latent period of studied races in Iranian wheat genotypes

Genome	TTTTF		PTRTF		TTKTK		TKTF	
	IT	LP	IT	LP	IT	LP	IT	LP
Original data								
Marker- trait association	29	15	15	18	10	19	15	10
Genome A	2	2	2	5	3	6	4	3
Genome B	17	8	9	6	3	2	8	5
Genome D	8	2	1	3	3	8	2	1
Unassembled Chromosomes	2	3	2	4	1	3	1	1
Imputed data								
Marker- trait association	208	128	109	106	83	160	104	60

Genome A	55	37	27	44	25	53	37	10
Genome B	95	69	67	35	39	75	54	33
Genome D	56	19	13	23	18	29	12	16
Unassembled Chromosomes	2	3	2	4	1	3	1	1

IT: Infection type, LP: Latent period

Table 8 shows the results of the Bonferroni correction at the 5% probability level of the association analysis of original SNPs. For Bonferroni correction of the original dataset, 0.05 was divided by the number of markers used (8959 markers) and finally were identified markers that were significant at the probability level of 5.58098E-06. Based on the results of Bonferroni correction, two QTLs for the TTTTF latent period, one QTL for the infection type and four QTLs for latent period of PTRTF, one QTL for the infection type, and one QTL for latent period of TTKTK, and one QTL for the infection type of TKTTF were identified. The identified QTLs associated with the latent period of the studied races, except for one QTL related to latent period of PTTTF that located on chromosome 7A with a distance of 133.92 cM, the rest were associated with markers whose genomic location was unknown. The QTL identified for the PTRTF infection type was located on chromosome 7A at 132.92 cM. QTL of TTKTK infection type was on chromosome 6B at 43.284 cM and QTL of TKTTF infection type was on chromosome 1D at 64.821 cM.

Bonferroni correction of the imputed SNPs was performed as with the original SNPs, but with this difference that, the number of markers used in this data was 43921 markers, so 0.05 was divided into these markers number (43921) and the level of significance was considered 1.1384E-06 and the results are shown in Table 9. In this regard, the TTTTF infection type was controlled by six QTLs, which one QTL on chromosome 1, two QTLs on chromosome 3 and one QTL on chromosome 5 of genome B was located; the other two QTLs were located on chromosome 2D. For the latent period of this race (TTTTF), there were 26 QTLs that eight QTLs was on the A genome, 12 QTLs was on the B genome and six QTLs was on the D genome. Chromosomes three, six and two of genome A, chromosomes one, six, three and two

of genome B and chromosomes four, seven and one of genome D carried QTLs identified in the TTTTF latent period. Regarding the infection type and latent period of PTRTF, one and 31 QTLs were observed, respectively. The only QTL of the infection type was present at 4.558 cM of chromosome 3A. from the 31 identified QTLs for latent period, three QTLs on chromosome 1A, three QTLs on chromosome 1B, one QTL on chromosome 1D, one QTL on chromosome 2A, one QTL on chromosome 2B, four QTLs on chromosome 3A, three QTLs were on chromosome 3B, three QTLs on chromosome 4D, six QTLs on chromosome 6A, one QTL on chromosome 6B and three QTLs were located on chromosome 7D; two QTLs were also associated with markers with unknown genomic loci. For the latent period of TTKTK, 14 QTL was detectable, but no QTL was recorded for the infection type. The identified QTLs of the latent period were located on A (chromosomes six, three, five and seven, respectively) and B genomes (chromosomes one and three, respectively) and a QTL was also associated with markers with unknown genomic loci. In the TKTTF, only three QTLs associated with the infection type were identified, two QTLs were located on chromosome 1B at 11.376 cM and one QTL was located on chromosome 1D at 64.821 cM.

Table 8. Most significant original SNP markers associated with quantitative trait loci for resistance to stem rust races of TTTTF, PTRTF, TTKTK, and TKTTF

Marker	Chr	Pos	P value							
			TTTTF		PTRTF		TTKTK		TKTTF	
			IT	LP	IT	LP	IT	LP	IT	LP
rs46629	1D	64.821	-	-	-	-	-	-	1.91E-6	-
rs23510	6B	43.284	-	-	-	-	3.56E-6	-	-	-
rs22132	7A	133.92	-	-	3.23E-6	2.81E-6	-	-	-	-
rs13588	Un	0	-	1.56E-6	-	3.38E-8	-	-	-	-
rs18901	Un	0	-	1.56E-6	-	3.38E-8	-	-	-	-
rs27487	Un	0	-	-	-	5.18E-6	-	2.75E-7	-	-

Chr: Chromosome, Pos: Position, IT: Infection type, LP: Latent period, Un: Unassembled Chromosome

Highly significant markers associated to infection type and latent period of *Pgt* races, their chromosomal sequence and position, the closest wheat gene or genes to them, orthologous genes (with the highest percentage of identity), identity percentage of wheat gene or genes that matches to the ortholog, molecular function and biological processes of the wheat gene or

genes associated with the markers and the extent of phenotype variance explain (R^2) are reported in table 10. Out of the 51 highly significant identified SNP markers, only 9 markers had the adjacent wheat gene or genes with the same chromosomal position. The Ensembl (<https://asia.ensembl.org/index.html>) site was used to gain further information about adjacent genes beyond the genetic position of the markers. The genes with the highest percentage of identity and the lowest E-value with significant markers were reported. The amount of phenotypic variance explained by SNP markers varied between 9 and 25%. The rs21674 and rs51316 markers on the 1B justified the high variance explained for the associated traits to other markers. These markers are in the closest genetic position to the TraesCS1B02G208400 and TraesCS1B02G037100 genes, respectively, and each of these genes have genetic identity with the orthologous genes of AET1Gv20503500 (96.74%) of *Aegilops tauschii* and TRIDC1AG003150 (95) of *Triticum dicoccoides*. The wheat TraesCS1D02G315800 gene, which with a rs46629 marker is located on the chromosome 1D, is in complete identity (100%) with the AET1Gv20751400 ortholog gene of *Aegilops tauschii*.

Table 9. Most significant imputed SNP markers associated with quantitative trait loci for resistance to stem rust races of TTTTF, PTRTF, TTKTK, and TKTTF

Marker	Chr	Pos	P value							
			TTTTF		PTRTF		TTKTK		TKTTF	
			IT	LP	IT	LP	IT	LP	IT	LP
rs63550	1A	44.512	-	-	-	1.15E-7	-	-	-	-
rs63551	1A	44.512	-	-	-	1.15E-7	-	-	-	-
rs63552	1A	44.512	-	-	-	1.15E-7	-	-	-	-
rs12408	1B	54.669	-	1.98E-7	-	-	-	-	-	-
rs21674	1B	58.08	-	3.47E-18	-	7.11E-19	-	1.35E-10	-	-
rs45874	1B	11.376	-	-	-	-	-	7.55E-7	2.11E-7	-
rs45875	1B	11.376	-	-	-	-	-	7.55E-7	2.11E-7	-
rs47978	1B	54.669	-	1.36E-7	-	2.69E-7	-	-	-	-
rs51316	1B	25.027	-	8.32E-9	-	1.12E-10	-	-	-	-
rs5923	1B	106.992	1.06E-6	-	-	-	-	-	-	-
rs33527	1D	47.767	-	3.89E-8	-	8.69E-8	-	-	-	-
rs46629	1D	64.821	-	-	-	-	-	-	5.21E-7	-
rs17477	2A	92.517	-	4.31E-8	-	1.31E-8	-	-	-	-
rs25281	2B	40.98	-	8.93E-7	-	2.62E-8	-	-	-	-
rs46339	2D	79.911	1.00E-6	-	-	-	-	-	-	-
rs46340	2D	79.911	1.00E-6	-	-	-	-	-	-	-
rs13329	3A	43.437	-	-	-	-	-	6.96E-8	-	-
rs50704	3A	4.558	-	5.03E-8	-	7.21E-8	-	-	-	-
rs50705	3A	4.558	-	5.03E-8	-	7.21E-8	-	-	-	-
rs50706	3A	4.558	-	5.03E-8	-	7.21E-8	-	-	-	-
rs62417	3A	4.558	-	-	-	-	-	1.68E-7	-	-
rs7318	3A	4.558	-	2.55E-10	6.70E-8	2.92E-13	-	-	-	-
rs20274	3B	31.882	5.93E-7	3.72E-18	-	1.77E-19	-	1.11E-10	-	-
rs20275	3B	31.882	5.93E-7	3.72E-18	-	1.77E-19	-	1.11E-10	-	-
rs20294	3B	31.882	-	2.75E-8	-	3.82E-9	-	-	-	-
rs33960	3B	62.576	-	-	-	-	-	1.78E-7	-	-
rs65252	4D	54.756	-	2.61E-8	-	7.93E-9	-	-	-	-
rs65253	4D	54.756	-	2.61E-8	-	7.93E-9	-	-	-	-
rs65254	4D	54.756	-	2.61E-8	-	7.93E-9	-	-	-	-
rs51204	5A	93.664	-	-	-	-	-	1.53E-7	-	-
rs22086	5B	86.610	4.63E-7	-	-	-	-	-	-	-
rs17868	6A	90.292	-	-	-	2.62E-7	-	-	-	-
rs34668	6A	50.208	-	2.56E-18	-	4.91E-19	-	1.91E-10	-	-
rs35063	6A	50.208	-	2.21E-18	-	4.78E-19	-	1.91E-10	-	-
rs35074	6A	50.208	-	2.74E-18	-	4.73E-19	-	1.95E-10	-	-
rs4500	6A	10.247	-	-	-	6.97E-7	-	-	-	-
rs53936	6A	99.391	-	-	-	1.80E-7	-	-	-	-
rs14828	6B	58.062	-	1.00E-6	-	-	-	-	-	-
rs14829	6B	58.062	-	1.00E-6	-	-	-	-	-	-
rs14830	6B	58.062	-	1.00E-6	-	-	-	-	-	-
rs25306	6B	47.831	-	-	-	6.48E-8	-	-	-	-
rs32944	6B	47.831	-	4.11E-8	-	-	-	-	-	-
rs48710	7A	71.904	-	-	-	-	-	8.54E-8	-	-
rs16240	7D	83.31	-	7.72E-9	-	4.73E-9	-	-	-	-
rs20930	7D	83.31	-	1.21E-8	-	3.81E-9	-	-	-	-
rs40699	7D	82.173	-	-	-	4.09E-7	-	-	-	-
rs13588	Un	0	-	-	-	3.38E-8	-	-	-	-
rs18901	Un	0	-	-	-	3.38E-8	-	-	-	-
rs27487	Un	0	-	-	-	-	-	2.75E-7	-	-

Chr: Chromosome, Pos: Position, IT: Infection type, LP: Latent period, Un: Unassembled Chromosome

Table 10. Basic local alignment search tool results of significant markers for seedling resistance to *Pgt* races

Marker	Sequence	Allele	Chromosome	Adjacent <i>T. aestivum</i> gene	Position (bp)	Orthologous gene	Identity	Molecular function	Biological process	Trait	R ²
rs4500	TGCAGACGGCGGGCGCGGAAC AAGCATGGTGGCGAAGCCGACA CGGAGGAGGATCTCGGCGAG	C/G	6A	TraesCS6A02G 100400	67,642,014- 67,642,038	TRIDC6A G012980 ^a	99.70	protein binding	-	PTRTF (LP)	0.09
rs12408	TGCAGCAAGCTAAACCAAGTCT TGAAACTGGAGCAGCGAACAAT TCTAAAGTCGCTTGAGGCTC	A/G	1B	TraesCS1B02G 330100	555,948,792- 555,952,831	TRIDC1B G053570 ^a	85.81	nucleotide binding, protein kinase activity, protein serine/threonine kinase activity, protein binding, ATP binding, kinase activity, transferase activity	protein phosphorylation , phosphorylation	TTTTF (LP)	0.10
rs21674	TGCAGCCAGAGCTCAAGAGGCC CCCTCGGCGGCATCTCAAAGGA GAAGGAAGGAGGAAGGCGAA	C/T	1B	TraesCS1B02G 208400	377,964,396- 377,970,306	AET1Gv2 0503500 ^b	96.74	DNA binding, lipid binding, sequence- specific DNA binding	regulation of transcription, DNA-templated	TTTTF (LP) PTRTF (LP) TTKTK (LP)	0.25 0.15 0.25
rs25281	TGCAGCCGCCCCTCCACCCCT TCGTCTTCGTCCAGCACCAACC CGTCGCCAGCGTCGCCGTCA	A/G	2B	TraesCS2B02G 269700	367,657,356- 367,709,290	AET2Gv2 0533600 ^b	96.94	protein binding	-	TTTTF (LP) PTRTF (LP)	0.09 0.11
rs33960	TGCAGCGTGTAGATTACCATGC TATTCACACAGAAGTTATCGGG TGTATATGTCATCAACTTTT	G/T	3B	TraesCS3B02G 260900	418,887,290- 418,898,340	TRIDC3B G039070 ^a	96.53	protein kinase activity, MAP kinase activity, ATP binding	protein phosphorylation	TTKTK (LP)	0.10
rs46339	TGCAGGCAGCTGACTGTACTGG AGGGCCACACATTTATCGAAAC CCTGCTGCCTGCCCTGGTTC	A/G	2D	TraesCS2D02G 465300	571,193,736- 571,202,194	TRIDC6A G061710 ^a	74.78	ADP binding	-	TTTTF (IT)	0.09
rs46340	TGCAGGCAGCTGACTGTACTGG AGGGCCACACATTTATCGAAAC CCTGCTGCCTGCCCTGGTTC	C/A	2D	TraesCS2D02G 465300	571,193,736- 571,202,194	TRIDC6A G061710 ^a	74.78	ADP binding	-	TTTTF (IT)	0.09
rs46629	TGCAGGCATTCCATCAAGCAAC CAGCAAGTCAGACCAGAATCA CCCCAATCAGACATGAGTGA	C/T	1D	TraesCS1D02G 315800	410,793,556- 410,798,235	AET1Gv2 0751400 ^b	100	serine-type endopeptidase activity, superoxide dismutase activity	protein processing, oxidation- reduction process	TKTTF (IT)	0.09
rs51316	TGCAGGGGGTAGCCCTGAGCGC CGCCAAATCAGATCAAACGATG CCCTTACTACATAAGCCAG	A/G	1B	TraesCS1B02G 037100	17,730,527- 17,739,883	TRIDC1A G003150 ^a	95.00	copper ion binding	defense response to bacterium	TTTTF (LP) PTRTF (LP)	0.12 0.15

a: *Triticum dicoccoides*, b: *Aegilops tauschii*, IT: Infetion type, LP: Latent period

Discussion

The characteristics of the races used to evaluate seedling stage and the effective and ineffective genes for each race are presented in Table 11. These races are known as the dominant black rust races in the country collected during 2015 and 2016 and with inoculated on 20 lines and differential cultivar of North American [20] the race of those have been determined. Isolate 95- 2 was pathogenic to the *Sr31* resistance gene, which for a very long time caused resistance to *Pgt*. This isolate belonged to the Ug99 race group and was named TTKTK.

The results of combined analysis of variance showed that in both traits of infection type and latent period, the highest total sum of squares was explained by genotype effect (56.55% and 49.79%, respectively); therefore, it is inferred that genotypes have had a significant effect on data variation. Then, the highest justified variance of infection type (34.40%) and latent period (32.98%) were due to genotype-race interaction. The race effect for infection type and latent period explained 3.75% and 7.85% of total variation, respectively. The small effect of race indicates a relatively low diversity among the races under study.

The values of the genetic and phenotypic variation coefficients for all races infection types and latent periods were very close to each other; this indicates the high impact of genes on creating diversity among genotypes. Each of the phenotypic, genetic, and environmental diversity coefficients provides useful information about the genetic or environmental diversity observed. As can be seen in the results (Table 3), the coefficient of phenotypic variation was higher than the genetic variation coefficient in all studied infection types and latent periods. This is due to the influence of environmental factors. Since the phenotypic variance is derived from the sum of environmental and genetic variance, so if the genetic variance is assumed to be constant, what causes one trait to differ in phenotypic and genetic variation would be environmental variance; In other words, if the trait has high phenotypic variation and low

genetic diversity, this indicates the effect of the environment. The greater the variation ratio of genetic to the environment, the greater the efficiency of selection, and the identification and selection of favorable genotypes from unfavorable will be more accurately performed.

Genetic variation coefficient reflects the variation among genotypes in terms of specificity under study and alone is not able to determine the extent of inheritance of this variation. This index, along with heritability, provides a good estimate of genetic progress in phenotypic selection [21]. Simultaneous application of two very important parameters of heritability and genetic efficiency plays an important role in the development of cultivars and genotypes. The high rate of genetic efficiency indicates additive gene action and its low level indicates non-additive gene action. When high heritability is estimated for a trait, genetic efficiency will not necessarily be high. If high heritability and high genetic efficiency coincide, it will reflect the additive effects of genes; but if high heritability is associated with low genetic function, it would indicate epistatic effects or dominance [22]. If trait heritability is less than 0.2 (20%), it indicates low heritability, if it is between 0.2 (20%) to 0.5 (50%) it has moderate heritability and if it is more than 0.5 (50%) it has high heritability [23]. High heritability with favorable genetic efficiency were observed in the PTRTF infection type and latent period (Table 3), confirming the additive effects of the gene, and indicating that a large proportion of phenotypic variation explains the genotypic variation. High inheritance along with high genetic efficiency is a very important factor in predicting the effects of selecting the best individuals in a population.

The results showed that in TTTTF, 16 cultivars (17.98%) and 34 landraces (16.34%), in PTRTF, 18 cultivars (20.22%) and 41 landraces (19.71%), in TTKTK, 34 cultivars (38.20%) and 65 landraces (31.25%) and in TKTTF, 16 cultivars (17.98) and 36 landraces (17.31%) had resistance reaction.

Genotypes were classified into two groups of resistant and susceptible genotypes according to the infection type (0-4 scale) at seedling stage. Difference in frequency of resistance reaction to races collected from wheat were evident in the germplasm. In the TTTTF, 16.77% of genotypes (16 cultivars and 34 landraces) showed resistance reaction and 83.23% of genotypes (248) had susceptible reaction. These genotypes are likely to have one or a combination of the effective resistance genes at the seedling stage of *Sr24* and *Sr31* that have caused a resistance reaction to this race. Concerning evaluation of seedling stage in relation to PTRTF, 20.22% of cultivars (18 cultivars) and 19.71% of landraces (41 landraces) showed resistance reaction and 71 cultivars (79.77%) and 167 landraces (80.29%) showed susceptible reaction. The genes that caused resistance to this race were *Sr21*, *Sr9e*, *Sr7b*, *Sr30*, *Sr24* and *Sr31*. This race (PTRTF) had lower virulence compared to the other races studied and break down fewer resistance genes number. The TTKTK race belong to the Ug99 race group and has caused many of the resistance genes identified to *Pgt* (*Sr5*, *Sr21*, *Sr9e*, *Sr7b*, *Sr11*, *Sr6*, *Sr8a*, *Sr9g*, *Sr9b*, *Sr30*, *Sr17*, *Sr9a*, *Sr9d*, *Sr10*, *SrTmp*, *Sr31*, *Sr38*, *SrMcN*) to be ineffective. One of these genes was the *Sr31* gene; it had for many years (over 40 years) maintained wheat genotypes and ultimately food security against the threat of *Pgt*, but with the emergence of the Ug99 race and its derived races, *Sr31* gene resistance was broken. Only two *Sr36* and *Sr24* genes has retained resistance to TTKTK. The presence of these genes either individually or in combination in 38.20% of cultivars (34 cultivars) and 31.25% of the landraces under study caused resistance to TTKTK. Absence of these genes in 61.80% of cultivars and 68.75% of landraces resulted in a susceptible reaction to TTKTK. Seedling stage investigation to TKTTF confirmed the presence of *Sr11*, *Sr24* and *Sr31* genes in 16 cultivars (17.98) and 36 landraces (17.31%) and their absence in 73 cultivars (82.02) and 36 landraces (82.69%) that the presence of these genes was followed the resistance reaction and their absence the susceptibility reaction. *Sr24* gene resistant was not broken by any of the races used, whereas

Sr31 gene resistance was only broken by TTKTK; However, with respect to the ineffectiveness of the *Sr31* gene in the TTKTK race, the number of resistance reaction genotypes was ranked first in terms of frequency of resistance reaction compared to other races; Therefore, it can be deduced that in the germplasm under evaluation, the presence of *Sr24* gene has caused relative resistance to all races of the stem rust disease. It should be noted, however, that the resistance generated may be caused by resistance genes not yet identified or their combination with identified effective resistance genes. Generally the majority of genotypes were susceptible to races which disagrees with previous studies [24-27].

Population structure has been used in genetic studies to explain relationship between individuals inside and between populations and shows a perspective on evolutionary relationship of individuals in a population. In addition existence of structure in population that studied for association mapping, it is a deterrent to achieving reliable result and caused on positive results in population that is not considered by the effects of population structure factors and relationship [28]. Out of 282 genotypes in both groups the original and imputed datasets, 277 genotypes (98.22%) belonged to the 3 identified subpopulations and the remaining 5 genotypes (1.78%), including 1 cultivar and 4 landraces, were excluded from the mixed genotypes. The number of these genotypes that were in mixed group, was small that indicating very low mixing in the population under study. Allelic incorporation in landraces of more than one single gene cohort could have caused a slight mixing, that due to the distribution of wheat in more than one ancestral population [29]. Another reason for mixing may be the presence of gene flow from the introduction of new genotypes into farms in the past. In this regard, germplasm exchange between different Mediterranean regions due to the development of ancient empires is considered as another possible reason for mixing [30]. Geographical origin of genotypes, selection and genetic drift cause subpopulation within a large population [31, 32]. The reason some cultivars fall into associated groups of landrace is

that some of cultivars have been selected and introduced based on indices from landrace and thus have a high genetic similarity to landrace [33].

The extent of linkage disequilibrium (LD) determines the number of markers needed to identify marker-trait association as well as mapping resolution [31]. On the other hand, in genome wide association mapping, locating QTLs is based on the extent of the LD and is therefore of particular importance [34]. Generally with increasing genetic distance, LD decreased. The range of LD varied across chromosomes as well as from chromosomes to other chromosomes. The share of each of the genomes A, B and D of the total original SNP markers in commercial cultivars and landraces was estimated to be approximately 37, 46 and 17%, respectively, and in the imputed SNP markers approximately 36, 50 and 13%, respectively. Nearly 87% of markers pairwise and 90% of significant markers pairwise in commercial cultivars and in landraces nearly 87% of markers pairwise and 88% of significant markers pairwise in original SNPs had genetic distance less than 10 cM. In the imputed SNPs, all the markers pairwise and the significant markers pairwise related to the commercial cultivars and landraces were located less than 10 cM. This issue indicates a high impact of LD on population, but as can be seen in result, with increasing distance between markers pairwise, LD in the population have decreased but not disappeared; this is due to other factors such as population structure, genetic drift, migration, selection and mutation. As can be seen in the results, the amount of LD in the cultivars was higher than landraces. The reason for this increase in LD in cultivars is the selection, whether natural or synthetic, which causes a LD among the genes selected and the genes associated with it. In addition, selection of the ascending (or descending) trait controlled by two or more non-linkage genes, despite the high genetic (physical) distance among the genes, increases the LD [35]. The relatively high level of LD in many chromosomal regions in the population indicates that association mapping can be a useful and effective method for identifying and confirming QTLs in these regions [36].

Since association mapping was introduced in plants [10], Due to significant advances in DNA sequencing technology, interest in the identification of new genes has gained popularity. The application of this method in plant populations to identify loci responsible for genetic variation including resistance to disease is increasing [37-39]. Knowledge of the genetic diversity of wheat in order to accelerate genetic and breeding studies is essential to produce disease resistant genotypes, because understanding the genetic nature of disease resistance can play a key role in the development of resistant genotypes. The use of genetic diversity of wheat germplasm to understand genetic mechanisms of disease resistance through GWAS can be very effective in breeding programs aimed at producing resistant varieties of *Pgt*.

Association mapping introduced a total of 69 QTLs for the infection type and 62 QTLs for the latent period in the original dataset and in the imputed dataset, 504 QTLs for the infection type and 454 QTLs for the latent period with a 0.1% probability level ($P \leq 0.001$). The highest number of QTLs identified for the infection type of studied races in the original and imputed dataset were non-uniformly distributed on the B genome chromosomes.

After Bonferroni correction aimed at identifying markers that are highly correlated with the target trait, the number of QTLs identified reduced in the original dataset, so the total of the QTLs for the studied races infection type were 3 and for the latent period 7 QTLs remained. In the imputed dataset, the number of QTLs of the infection type was reduced to 10 QTLs and the number of QTLs of the latent period decreased to 71 QTLs. To answer the question of whether QTLs identified with specific genomic regions in this study have been previously introduced or not? The results of this study were compared with previous findings. Some of the QTLs identified in this study corresponded to the genomic regions of the *Sr* genes and the QTLs reported in previous studies [40, 41]. B chromosomes were found chromosomes carrying rust resistance genes in wheat, which has been reported in various studies [19, 27, 40, 42], in the present study, only one marker (rs23510) was identified based on the results of

genome B located on chromosome 6 and was associated with the genomic locus that caused resistance to TTKTK. The number of markers in genome B associated with the infection type and the latent period were significantly higher in the imputed dataset than in the original dataset and were distributed on chromosomes 1B, 6B, 3B, 2B and 5B, respectively. Letta *et al.* (2014) identified a QTL on chromosome 1B for resistance to the JRCQC; this disease-resistant genomic region has also been mentioned. The rs51316 marker associated with the TTTTF and PTRTF infection types was largely close to the QTL that identified by Letta *et al.* On the other hand, in this chromosomal position, the *Sr14* gene is resistant to some *Pgt* races in wheat [43], but in this study a significant effect on the infection type of each race was observed. Specific study of TTKTK infection type belonging to the Ug99 group and reported in the study of Jin *et al.* [44] did not show any effect. At 106.99 cM is located on chromosome 1B marker rs5923. This marker is related to the genomic region that caused seedling resistance to the TTTTF and was reported in the study by Letta *et al.* [19]. Not only the *Pgt* resistance gene is located in this region, but also the Lr46/Yr29/Pm39 gene block and the unknown resistance genes are in the adult plant stage [45, 46]. A number of *Pgt* resistance genes, including *Sr9*, *Sr16*, and *Sr28* [18] and *SrWeb*, are located on the long arm of chromosome 2B. The *SrWeb* gene causes resistance to Ug99, while none of the four *Sr9* alleles has the ability [44]. No resistance to TTKTK was found on chromosome 2B; the *SrWeb* gene may not be present in the germplasm under study or there is no marker associated with this gene. After genome B, genome A had the highest marker-trait association. In the original dataset after Bonferroni correction, only one marker (rs22132) was identified in the A genome, located on chromosome 7 at 133.92 cM and was associated with both PTRTF infection type and latent period characteristics. There were no QTLs in genome A in the imputed SNPs for the infection type of any of the race studied., but for the latent period, 8 QTL (1 QTL on 2A, 4 QTL on 3A and 3 QTL on 6A) in TTTTF, 14 QTL (3 QTL on 1A, 1

QTL on 2A, 4 QTL on 3A and 6 QTL on 6A) in PTRTF and 7 QTL (2 QTL on 3A, 1 QTL on 5A, 3 QTL on 6A and 1 QTL on 7a) was identified in TTKTK. TKTTF was not associated with genomic A chromosomes. The *Sr34* and *Sr38 Pgt* genes are located on chromosome 2A [19] and are somewhat close to the rs17477 marker in genomic position, but because the resistance of these two genes is broken by the studied races and as a result, they were inactivated against rust disease, with no QTL-associated region associated with the infection type. Therefore, this marker can be used in marker assisted selection to identify *Sr34* and *Sr38*. In a study, the coexistence of two markers cfa2201 and wPt-5839 with two *Sr34* and *Sr38 Pgt* resistance genes was reported [19]. These two genes are ineffective against the Ug99 race group, one of which (TTKTK) has been investigated in the present study (Jin *et al.*, 2007; Singh *et al.*, 2011). The *Sr7* and *SrND643* resistance genes (probably the *Sr7* allele) are located on the long arm of chromosome 4A [18, 47]. This gene caused resistance to the PTRTF race under study, but no marker-trait association was found in this study. Other studies have also mapped genomic regions associated with resistance to black rust disease in genome A such as marker wpt-734078 [45] and wpt-6869 [48] on chromosome 1AS.

The D genome had the lowest number of marker-trait association. In the original dataset between the infection type and the latent period of all the races under study, only one QTL was identified for the TKTTF infection type associated with the rs46629 marker on the 1D chromosome. In the imputed dataset, this marker retained its association with the TKTTF infection type, and the only marker had the marker-trait association on all TKTTF D genome chromosomes. The TTKTK had no marker-trait association with the D genome. For the TTTTF infection type, 2 QTLs were on the 2D and for the latent period 1 QTL was on 1D, 3 QTLs on 4D and 2 QTLs on 7D. In PTRTF, 1 QTL was observed on 1D, 3 QTL on 4D and 3 QTL on 7D. Edae *et al.* [49] in their study reported QTHJC resistance-related IWB17135 marker and TPMKC resistance-related IWA2415 marker on the short arm of 2D chromosome.

Tsilo *et al.* [50] mapped a QTL on the 2D chromosome at 78.50 cM, which was 1.1 cm from the Sr6 *Pgt* resistance gene.

In the original dataset, there were 6 QTLs on unknown chromosomes, 2 QTLs for TTTTF latent period, 3 QTLs for PTRTF latent period and 1 QTL for TTKTK latent period. In the imputed dataset, 2 QTLs associated to PTRTF latent period and 1 QTL associated to TTKTK latent period that overlapped with the original dataset were identified.

The markers that had the highest marker-trait association were blasted at the Ensembl database to identify overlapping genes, their molecular function and biological processes, as well as to identify orthologous genes. The results indicate that the genes adjacent to the markers play an important role in biosynthetic pathways such as ion transport, oxidation reduction, protein processing, phosphorylation, and so on. In general, environmental stresses cause significant changes in the expression levels of many of these genes in plants, so such changes result in the accumulation or reduction of important metabolites, changes in enzyme activity and protein synthesis rates, as well as the production of new proteins [51] that deal with environmental stress in various ways. ROS, phospholipid-derived signals, and cyclic nucleotide-dependent signals, as well as some plant hormones, are involved in stress signaling [52, 53]. When the plant is attacked by the pathogen, in response to the pathogen attack, ROS is produced in the plants and acts as an important sign of stress and reduces the amount of stress damage possible by activating defense mechanisms [52-54]. Membrane lipids by regulating membrane fluidity and other physicochemical properties cause secondary signaling molecules in response to stress. Biosynthesis of lipids and their degrading enzymes play several roles such as direct or indirect regulation or effect on stress signaling and tolerance [52]. In plants, a variety of biotic and abiotic stresses such as pathogens, drought, salinity, high temperature, intense light, hormones, and nodulation factors cause changes in cytosolic calcium levels and lead to stress transmission [55, 56]. In many transcriptional signaling

pathways, the major form of signal transduction is reversible protein phosphorylation. Protein kinases such as mitogen-activated protein kinase (MAPK) are essential in stress developmental, hormonal, biotic and abiotic signaling. Protein phosphatases are responsible for the dephosphorylating of phospho-proteins. Protein phosphatases are subdivided into serine/threonine phosphatase, tyrosine phosphatase, and dual phosphatases through their specific substrates, which play an important role in serine/threonine phosphatase transcriptional stress transcription [52]. Eventually, protein kinases are likely to target transcription factors and bind to the stress-responsive genes promoter and consequently activate transcription [52].

After identifying the nearest wheat gene or genes to the markers, orthologous genes with these genes were examined. Numerous orthologous genes were observed but only genes with high percentage of identity were selected. The orthologous genes identified were generally related to two species of *Triticum dicoccoides*, *Aegilops tauschii*, ancestors of wheat and during the evolutionary stages of wheat with hybridization, they have entered it [57-62]. These orthologous genes had a high percent of identity with the genes identified in wheat that are involved in stress resistance mechanisms.

Conclusion

Association mapping of resistance of Iranian bread wheat cultivars and landraces to four *Pgt* races in seedling stage in terms of latent period and infection type. The studied germplasm was grouped into three sub-populations. Based on the results of original and imputed dataset, the highest number of marker pairwise in cultivars and landraces was observed on B genomes. In the original dataset, the highest linkage disequilibrium was observed in D genome of cultivars and landraces but in imputed dataset, was observed in genome A of cultivars and landraces. The results showed a significant variation in the amount of linkage disequilibrium

on different chromosomes. However, the cultivars had higher linkage disequilibrium compared to the landraces, which could be attributed to the selection made during the production and release stages of the cultivars. Genome wide association study based on original dataset revealed 69 and 62 marker-trait association for races infection type and their latent period, respectively. The number of marker-trait association in the imputed dataset was 504 for infection type and 454 for the latent period. When Bonferroni correction was performed, in the original dataset, two QTLs for the TTTTF latent period, one QTL for the infection type and four QTLs for the latent period of PTRTF, one QTL for the infection type, and one QTL for the latent period of TTKTK, and one QTL for the TKTTF infection type remained. In the alternative data, the TTTTF infection type by six QTLs and its latent period by 26 QTLs, the PTRTF infection type and latent period by one and 31 QTLs respectively, the TTKTK latent period by 14 QTLs, and the TKTTF infection type by three QTLs was controlled. No marker-trait association was detected in TTKTK infection type and TKTTF latent period. Nine markers out of 51 SNP markers (very significant) had a chromosomal position similar to the adjacent *Triticum aestivum* gene or genes. By studying the molecular function and biological processes of these genes, it was found that these genes have different mechanisms that cause stress resistance. The results obtained from this study can be important to facilitate and accelerate breeding programs through marker-assisted selection.

Methods

Plant materials

A set of 297 Iranian bread wheat genotypes, including 208 landraces collected between 1931 to 1968 and 89 commercial cultivars released between 1942 to 2014 along with the susceptible cultivar Morocco were tested for phenotypic evaluation but genotypic data of 282 genotypes were available, thus AM was performed based on 282 genotypes. Landraces were

obtained from the Gene Bank of the University of Tehran and commercial cultivars from Seed and Plant Improvement Institute (SPII), Karaj, Alborz, Iran and Dryland Agricultural Research Institute (DARI), Maragheh, East Azarbaijan, Iran. A detailed description of the genotypes at the molecular and phenotypic level is attached as the additional file 1, Table S1.

Phenotypic evaluation

Phenotypic evaluation of wheat genotypes against *Pgt* was carried out at the Cereal Research Department, Seed and Plant Improvement Institute (SPII), Karaj, Alborz, Iran in 2019. Twelve to 15 seeds of each genotype were planted in pots with 10 cm height and diameter containing 1:1:2 ratio of common soil, peat moss and leaf mold. Four wheat genotypes were placed in each pot at an appropriate distance. Then they were stored at 22-25 °C for development. After 8-10 days, when the first leaf of the seedlings developed well, the spores of four stem rust race collected from different regions of Iran (Table 11) were inoculated separately. The inoculated plants were kept in a dark chamber for 24 h at 18±2 °C and near saturation moisture, then transferred to a 22±2 °C greenhouse with a 16 h photoperiod. Inoculated plants were examined from the second day to record the latent period (LP). Fourteen days after inoculation, seedling infection type (IT) was recorded using the 0-4 scale introduced by Stakman *et al.* [17] and modified by McIntosh *et al.* [18]. To ensure the accuracy of the data collected, two replicates were considered for plant materials inoculated with each isolate. ITs less than or equal to 2+ were considered low infection types whereas ITs greater than or equal to 3 were considered high infection types. In order to use the modified Stakman ITs in the genome-wide association studies (GWAS), the 0-4 scale was converted to a 1-13 linear. The average linear scale score across the two replications was used in the AM analyses.

Table 11. Detailed description of the *Pgt* races used to evaluate the wheat genotypes

Isolate	Location	Race	Stem rust resistance (<i>Sr</i>) genes	
			Ineffective	Effective
94- 8	Borojerd, Lorestan, Iran	TTTTF	<i>Sr5, Sr21, Sr9e, Sr7b, Sr11, Sr6, Sr8a, Sr9g, Sr36, Sr9b, Sr30, Sr17, Sr9a, Sr9d, Sr10, SrTmp, Sr38, SrMcN</i>	<i>Sr24, Sr31</i>
94- 15	Kelardasht, Mazandaran, Iran	PTRTF	<i>Sr5, Sr11, Sr6, Sr8a, Sr9g, Sr36, Sr9b, Sr17, Sr9a, Sr9d, Sr10, SrTmp, Sr38, SrMcN</i>	<i>Sr21, Sr9e, Sr7b, Sr30, Sr24, Sr31</i>
95- 2	Shavour, Khouzestan, Iran	TTKTK	<i>Sr5, Sr21, Sr9e, Sr7b, Sr11, Sr6, Sr8a, Sr9g, Sr9b, Sr30, Sr17, Sr9a, Sr9d, Sr10, SrTmp, Sr31, Sr38, SrMcN</i>	<i>Sr36, Sr24</i>
95- 31	Kelardasht, Mazandaran, Iran	TKTTF	<i>Sr5, Sr21, Sr9e, Sr7b, Sr6, Sr8a, Sr9g, Sr36, Sr9b, Sr30, Sr17, Sr9a, Sr9d, Sr10, SrTmp, Sr38, SrMcN</i>	<i>Sr11, Sr24, Sr31</i>

Phenotypic data analysis

Before performing the analysis of variance, the assumptions of variance analysis were examined and the results confirmed the assumptions. To calculate genetic parameters, analysis of variance was performed for the infection type and latent period of the studied *Pgt* races. Genetic, environmental, and phenotypic variances were calculated using the Comstock

and Robinson [63] method, using equations $\sigma_g^2 = \frac{MS_g - MS_e}{r}$, $\sigma_e^2 = MS_e$ and $\sigma_p^2 = \sigma_g^2 + MS_e$, respectively. In these equations MS_g is genotype mean square, MS_e is mean square error and r

is the number of experimental replications. Heritability was estimated by the Falconer [64] method through the $h^2 = \frac{\sigma_g^2}{\sigma_p^2}$ equation where σ_g^2 is the genetic variance and σ_p^2 is the

phenotypic variance obtained from the variance analysis table according to Comstock and Robinson [63] method. To estimate the coefficient of genetic and phenotypic variation, Singh

and Chaudhary [65] method was used based on $GCV (\%) = \frac{\sqrt{\sigma_g^2}}{\bar{x}} \times 100$ and $PCV (\%) =$

$\frac{\sqrt{\sigma_p^2}}{\bar{x}} \times 100$ equations, where \bar{x} is the average trait. Genetic efficiency were also calculated

according to the Allard [66] and Singh and Chaudhry [65] method using the $GA = k \times \sigma_p \times h^2$ equation in which k is 10% of the selection pressure (1.75), σ_p is the phenotypic

standard deviation and h^2 is the heritability.

Simple and combined analysis of variance (ANOVA) on the linearized 1 to 13 ratings was performed using SAS software 9.3 [67] and genetic parameters were calculated using Excel.

The races grouping were performed according to the estimated infestation type of 0-4 scale (1-13 ratings) on Ward's method using SPSS software. Cluster count (cut line) was also determined based on the discriminant function.

Genotyping by sequencing and imputation method

Genotypic evaluation of wheat genotypes were performed in collaboration with the US Ministry of Agriculture and the University of Kansas. The genomic library was constructed according to the method of Poland *et al.* [68]. To do this, DNA was extracted, normalized, and digested with *PstI* and *MspI* enzymes. Then, barcode adapters were attached to each sample, so that each sample has a different barcode. To remove extras except barcoded genomic DNA, purification was performed by QIAquick (Qiagen) PCR purification kits. Finally, the size of the amplified fragments between 250-300 bp was determined on the E-gel system and sent for sequencing by the Ion Proton system. Sequencing data were treated for 64 bp and the same reads were grouped into tags. Identical sequence tags were aligned to identify the SNPs within the tags and the SNPs were summoned using the UNEAK (Universal Network Enabled Analysis Kit) GBS pipeline [69] as part of the TASSEL 4.0 bioinformatics package [70]. To reduce false positive error results, SNPs with heterozygosity greater than 10%, minor allele frequency less than 5%, and missing data greater than 20% were eliminated. After specifying the haplotype phase for all individuals, the data was subjected to imputation using BEAGLE v3.3.2 [71] based on available allele frequencies obtained. During imputation, four different reference genomes were assessed that among them W7984 reference genome was shown to have the greatest imputation accuracy [72]. The different

chromosomes LD decay was obtained using the `ggplot2` package in RStudio [73] based on LOESS regression.

Population structure and kinship matrix

In order to determine the genetic structure of the population of wheat genotypes and to identify sub-populations, Bayesian methods using STRUCTURE v2.3.4 [74] was used. The optimal number of K was determined by an admixture model and with a burn-in and simulation phase consisting of 30,000 steps for values of K from 1 to 10. The ΔK statistic was used to determine the most desirable subpopulation number and its plot was plotted for consecutive K values. The observed and expected values allele frequencies were used to estimate the linkage disequilibrium between markers in TASSEL v.5 [70]. Then, population structure matrix Q ($n \times p$), where n is the number of genotypes assayed and p is the number of subpopulations defined, was used in association studies. Cluster analysis and principal component analysis were also performed to determine the genetic relationships between the genotypes.

Genome-wide association study

Both general linear model (GLM) and mixed linear model (MLM) were used to examine the accurate association between marker and trait in TASSEL [70]. The GAPIT package [75] was also used to perform association mapping in both GLM and MLM methods in RStudio [73]. The results of TASSEL and GAPIT were investigated using t-test. Based on the results, it was found that the general linear model derived from TASSEL provides more accurate information about the marker-trait association.

Gene annotation

Sequences surrounding of significant SNP markers were obtained from the wheat 90 K SNP database [76] and used for assessing gene annotation using Gramene (<http://www.gramene.org/>) by aligning them to the IWGSC RefSeq v1.0 annotation (<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations>). The function of putative genes was explored by investigating the pathways which the encoded enzymes were involved in. After aligning SNPs sequences to the reference genome, overlapping genes with the highest identity percentage and blast score were selected for further processing. The ontology of each adjacent genes with *T. aestivum*, including molecular function and biological process and also orthologous genes in related species, were extracted from the ensemble-gramene database (<http://ensembl.gramene.org>).

Supplementary information

Additional file 1 Table S1. The information of Iranian wheat genotypes including cultivars released between 1942 and 2014, and landraces collected between 1931 and 1968

Abbreviations

Pgt: *Puccinia graminis* f. sp. *tritici*, *Sr*: Stem rust, QTL: Quantitative trait loci, AM: Association mapping, LP: Latent period; IT: Infection type; GBS: Genotyping by sequencing; GWAS: Genome-wide association study; LD: Linkage disequilibrium; MAF: Minor allele frequency; NSP: Number of SNP pairwise, SSP: Significant SNP pairwise, MAS: Marker-assisted selection; MTA: Marker-trait associations; SNP: Single nucleotide polymorphism, Chr: Chromosome, Pos: Position, Un: Unassembled Chromosome.

Acknowledgments

Thanks and appreciation to the Seed and plant improvement institute for making this research possible.

Authors' contributions

ASR: Performed greenhouse trial, analyzed the datasets, and wrote the manuscript. MRB: Assistance in data analysis and writing the manuscript. AM: Assistance in revising the manuscript. KM and HA: Assistance in data analysis and writing the manuscript. All authors contributed to revising and editing the manuscript. All authors have read and approved of the final manuscript.

Funding

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. The funding body was involved in the material creation, designing the study, data analysis and writing the manuscript.

Reference

1. FAO: **World food and agriculture** 2017.
2. statistics MoAJ: **Agriculture-Iran-Statistics**. In.; 2017.

3. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C: **Food security: the challenge of feeding 9 billion people.** *science* 2010, **327**(5967):812-818.
4. Oerke EC: **Crop losses to pests.** *Journal of Agricultural Science* 2006, **144**:31-43.
5. Schafer J, Roelfs A, Bushnell W: **Contributions of early scientists to knowledge of cereal rusts.** *The cereal rusts* 1984, **2**:3-38.
6. Roelfs AP: **Rust diseases of wheat: concepts and methods of disease management:** Cimmyt; 1992.
7. Singh RP, Hodson DP, Jin Y, Lagudah ES, Ayliffe MA, Bhavani S, Rouse MN, Pretorius ZA, Szabo LJ, Huerta-Espino J: **Emergence and spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control.** *Phytopathology* 2015, **105**(7):872-884.
8. Holland JB: **Genetic architecture of complex traits in plants.** *Current Opinion in Plant Biology* 2007, **10**:156-161.
9. Gupta SK, Charpe A, Koul S, Prabhu KV, Haq QMR: **Development and validation of molecular markers linked to an *Aegilops umbellulata*-derived leaf-rust-resistance gene, Lr9, for marker-assisted selection in bread wheat.** *Genome* 2005, **48**(5):823-830.
10. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: **Dwarf8 polymorphisms associate with variation in flowering time.** *Nature genetics* 2001, **28**(3):286-289.
11. Rafalski A, Morgante M: **Corn and humans: recombination and linkage disequilibrium in two genomes of similar size.** *TRENDS in Genetics* 2004, **20**(2):103-111.
12. Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF: **Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties.** *Proceedings of the National Academy of Sciences* 2006, **103**(49):18656-18661.
13. Kraakman ATW, Niks RE, Van den Berg PMM, Stam P, Van Eeuwijk FA: **Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars.** *Genetics* 2004, **168**:435-446.
14. Pasam RK, Sharma R, Malosetti M, Eeuwijk F, Haseneyer G, Kilian B, A. G: **Genome wide association studies for agronomical traits in a worldwide spring barely collection.** *BMC Plant Biology* 2012, **12**:1-16.
15. Moose SP, Mumm RH: **Molecular plant breeding as the foundation for 21st century crop improvement.** *Plant Physiology* 2008, **147**:969-977.
16. Bartlett M: **The general canonical correlation distribution.** *The Annals of Mathematical Statistics* 1947:1-17.
17. Stakman EC, Stewart D, Loegering W: **Identification of physiologic races of *Puccinia graminis* var. *tritici*:** USDA Washington; 1962.
18. McIntosh RA, Wellings CR, Park RF: **Wheat Rusts: An Atlas of Resistance Genes.** : CSIRO Publications, Victoria, Australia; 1995.
19. Letta T, Olivera P, Maccaferri M, Jin Y, Ammar K, Badebo A: **Association mapping reveals novel stem rust resistance loci in durum wheat at the seedling stage.** *Plant Genome* 2014, **7**.
20. Jin Y, Szabo LJ, Pretorius Z, Singh R, Ward R, Fetch Jr T: **Detection of virulence to resistance gene Sr24 within race TTKS of *Puccinia graminis* f. sp. *tritici*.** *Plant Disease* 2008, **92**(6):923-926.
21. Burton GW, Devane dE: **Estimating heritability in tall fescue (*Festuca arundinacea*) from replicated clonal material 1.** *Agronomy Journal* 1953, **45**(10):478-481.
22. Johnson HW, Robinson H, Comstock R: **Estimates of genetic and environmental variability in soybeans 1.** *Agronomy journal* 1955, **47**(7):314-318.
23. Stansfield W: **Molecular Genetics. Theory and Problems of Genetics, 3rd ed, Schaum's Outline Series, McGraw-Hill** 1991:356.

24. Singh R: **Pathogenicity variations of Puccinia recondita f. sp. tritici and P. graminis f. sp. tritici in wheat-growing areas of Mexico during 1988 and 1989.** *Plant disease* 1991, **75**(8):790-794.
25. Huerta-Espino J, Roelfs A: **Leaf rust on durum wheats.** *Vortraege fuer Pflanzenzuechtung (Germany)* 1992.
26. Ordoñez M, Kolmer J: **Virulence phenotypes of a worldwide collection of Puccinia triticina from durum wheat.** *Phytopathology* 2007, **97**(3):344-351.
27. Aoun M, Breiland M, Kathryn Turner M, Loladze A, Chao S, Xu SS, Ammar K, Anderson JA, Kolmer JA, Acevedo M: **Genome-wide association mapping of leaf rust response in a durum wheat worldwide germplasm collection.** *The plant genome* 2016, **9**(3).
28. Breseghello F, Sorrells ME: **Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars.** *Genetics* 2006, **172**(2):1165-1177.
29. Oliveira HR, Campana MG, Jones H, Hunt HV, Leigh F, Redhouse DI, Lister DL, Jones MK: **Tetraploid wheat landraces in the Mediterranean basin: taxonomy, evolution and genetic diversity.** *Plos one* 2012, **7**(5).
30. Moragues M, Moralejo M, Sorrells ME, Royo C: **Dispersal of durum wheat [*Triticum turgidum* L. ssp. turgidum convar. durum (Desf.) MacKey] landraces across the Mediterranean basin assessed by AFLPs and microsatellites.** *Genetic resources and crop evolution* 2007, **54**(5):1133-1144.
31. Flint-Garcia SA, Thornsberry JM, Buckler IV ES: **Structure of linkage disequilibrium in plants.** *Annual review of plant biology* 2003, **54**(1):357-374.
32. Buckler IV ES, Thornsberry JM: **Plant molecular diversity and applications to genomics.** *Current opinion in plant biology* 2002, **5**(2):107-111.
33. Alipour H, Bihanta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G: **Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars.** *Frontiers in plant science* 2017, **8**:1293.
34. Al-Maskri AY, Sajjad M, Khan SH: **Association Mapping: A Step Forward to Discovering New Alleles for Crop Improvement.** *International Journal of Agriculture & Biology* 2012, **14**(1).
35. Ataei R, Gholamhoseini M, Mohammadi V: **Association mapping in plants.** *Crop Biotechnology* 2017, **18**:28-37.
36. Zhang D, Bai G, Zhu C, Yu J, Carver BF: **Genetic diversity, population structure, and linkage disequilibrium in US elite winter wheat.** *The Plant Genome* 2010, **3**(2):117-127.
37. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R: **Linkage disequilibrium in the human genome.** *Nature* 2001, **411**(6834):199-204.
38. Hall D, Tegström C, Ingvarsson PK: **Using association mapping to dissect the genetic basis of complex traits in plants.** *Briefings in functional genomics* 2010, **9**(2):157-165.
39. Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, Hinze M, Plieske J, Kulosa D, Ganai MW: **Whole genome association mapping of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.).** *PLoS One* 2013, **8**(2).
40. Yu L-X, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME: **Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm.** *Theoretical and Applied Genetics* 2011, **123**(8):1257-1268.
41. Yu L-X, Morgounov A, Wanyera R, Keser M, Singh SK, Sorrells M: **Identification of Ug99 stem rust resistance loci in winter wheat germplasm using genome-wide association analysis.** *Theoretical and Applied Genetics* 2012, **125**(4):749-758.
42. Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J: **Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure.** *Genetics* 2007, **177**(3):1889-1913.

43. Singh RP, Hodson DP, Jin Y, Huerta-Espino J, Kinyua MG, Wanyera R, Njau P, Ward RW: **Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen.** *CAB reviews: perspectives in agriculture, veterinary science, nutrition and natural resources* 2006, **1**(54):1-13.
44. Jin Y, Singh R, Ward R, Wanyera R, Kinyua M, Njau P, Fetch T, Pretorius Z, Yahyaoui A: **Characterization of seedling infection types and adult plant infection responses of monogenic Sr gene lines to race TTKS of Puccinia graminis f. sp. tritici.** *Plant Disease* 2007, **91**(9):1096-1099.
45. Bhavani S, Singh R, Argillier O, Huerta-Espino J, Singh S, Njau P, Brun S, Lacam S, Desmouceaux N, McIntosh R: **Mapping durable adult plant stem rust resistance to the race Ug99 group in six CIMMYT wheats.** In: *Borlaug Global Rust Initiative 2011 Technical Workshop June 13-16 Saint Paul, Minnesota, USA: 2011.* 42.
46. Singh A, Pandey M, Singh A, Knox R, Ammar K, Clarke J, Clarke F, Singh R, Pozniak C, DePauw R: **Identification and mapping of leaf, stem and stripe rust resistance quantitative trait loci and their interactions in durum wheat.** *Molecular breeding* 2013, **31**(2):405-418.
47. Basnet BR, Singh S, Lopez-Vera EE, Huerta-Espino J, Bhavani S, Jin Y, Rouse MN, Singh RP: **Molecular mapping and validation of SrND643: A new wheat gene for resistance to the stem rust pathogen Ug99 race group.** *Phytopathology* 2015, **105**(4):470-476.
48. Rouse MN, Talbert LE, Singh D, Sherman JD: **Complementary epistasis involving Sr12 explains adult plant resistance to stem rust in Thatcher wheat (*Triticum aestivum* L.).** *Theoretical and Applied Genetics* 2014, **127**(7):1549-1559.
49. Edae EA, Pumphrey MO, Rouse MN: **A genome-wide association study of field and seedling response to individual stem rust pathogen races reveals combinations of race-specific genes in North American spring wheat.** *Frontiers in plant science* 2018, **9**:52.
50. Tsilo TJ, Chao S, Jin Y, Anderson JA: **Identification and validation of SSR markers linked to the stem rust resistance gene Sr6 on the short arm of chromosome 2D in wheat.** *Theoretical and applied genetics* 2009, **118**(3):515-524.
51. Zhu J-K: **Abiotic stress signaling and responses in plants.** *Cell* 2016, **167**(2):313-324.
52. Jamei R, Kargar-Khorami S, Zanganeh R: **Plant cellular and molecular mechanism against biotic stress.** In.: Urmia University Press; 2018: 290.
53. Ali M, Cheng Z, Ahmad H, Hayat S: **Reactive oxygen species (ROS) as defenses against a broad range of plant fungal infections and case study on ROS employed by crops against Verticillium dahliae wilts.** *Journal of plant interactions* 2018, **13**(1):353-363.
54. Lehmann S, Serrano M, L'Haridon F, Tjamos SE, Metraux J-P: **Reactive oxygen species and plant resistance to fungal pathogens.** *Phytochemistry* 2015, **112**:54-62.
55. Pandey GK, Kanwar P, Singh A, Steinhorst L, Pandey A, Yadav AK, Tokas I, Sanyal SK, Kim B-G, Lee S-C: **Calcineurin B-like protein-interacting protein kinase CIPK21 regulates osmotic and salt stress responses in Arabidopsis.** *Plant physiology* 2015, **169**(1):780-792.
56. Monihan S: **The Arabidopsis Calcineurin B-Like10 Calcium Sensor Couples Environmental Signals to Developmental Responses.** 2011.
57. Kihara H: **Discovery of the DD-analyzer, one of the analyzers of Triticum vulgare.** *Agriculture and Horticulture* 1944, **19**:889-890.
58. McFadden ES, Sears ER: **The origin of Triticum spelta and its free-threshing hexaploid relatives.** *Journal of Heredity* 1946, **37**(3):81-89.
59. Kerber E: **Wheat: reconstitution of the tetraploid component (AABB) of hexaploids.** *Science* 1964, **143**(3603):253-255.
60. Kislev ME: **TRITICUM PARVICOCUM SP. NOV., THE OLDEST NAKED WHEAT.** *Israel Journal of Botany* 1979, **28**(2):95-107.
61. Dvorak J, Luo M-C, Yang Z-L, Zhang H-B: **The structure of the Aegilops tauschii gene pool and the evolution of hexaploid wheat.** *Theoretical and Applied Genetics* 1998, **97**(4):657-670.

62. Matsuoka Y, Nasuda S: **Durum wheat as a candidate for the unknown female progenitor of bread wheat: an empirical study with a highly fertile F 1 hybrid with *Aegilops tauschii* Coss.** *Theoretical and Applied Genetics* 2004, **109**(8):1710-1717.
63. Comstock R, Robinson H: **Genetic parameters, their estimation and significance.** In: *Proceedings of the 6th international Grassland congress: 1952.* 248-291.
64. Falconer D: **Introduction to Quantitative Genetics (3rd edn, 438 p.).** In.: New York: Longmans Halow, Longman Sci e Tech; 1989.
65. Singh RK, Chaudhary BD: **Biometrical methods in quantitative analysis:** Kalyani, Publishers. New Delhi New Delhi, India; 1985.
66. Allard R: **Principies of Plant Breeding John Wiley.** New York 1960.
67. Guide As: **SAS® 9.3 In-Database Products.** 2011.
68. Poland JA, Brown PJ, Sorrells ME, Jannink JL: **Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach.** *PloS one* 2012, **7**(2).
69. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE: **Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol.** *PloS one* 2013, **9**(1).
70. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: **TASSEL: software for association mapping of complex traits in diverse samples.** *Bioinformatics* 2007, **23**(19):2633-2635.
71. Browning BL, Browning SR: **A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals.** *The American Journal of Human Genetics* 2009, **84**(2):210-223.
72. Alipour H, Bai G, Zhang G, Bihamta MR, Mohammadi V, Peyghambari SA: **Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references.** *PloS one* 2019, **14**(1).
73. Team R: **RStudio: integrated development for R.** *RStudio, Inc, Boston, MA* URL <http://www.rstudio.com> 2015, **42**:14.
74. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**(2):945-959.
75. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: **GAPIT: genome association and prediction integrated tool.** *Bioinformatics* 2012, **28**(18):2397-2399.
76. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L: **Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array.** *Plant biotechnology journal* 2014, **12**(6):787-796.

Figures

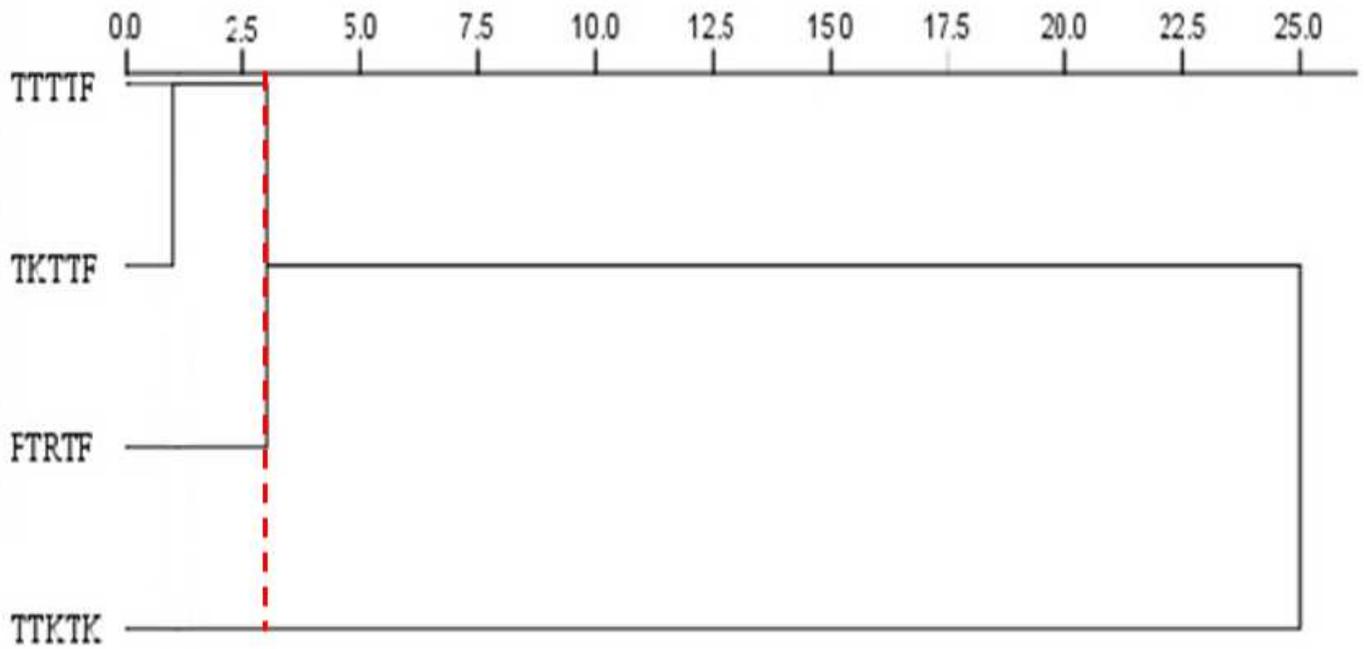


Figure 1

Dendrogram of cluster analysis of four Pgt races based on infection types and latent periods of 297 wheat genotypes using the Ward's method

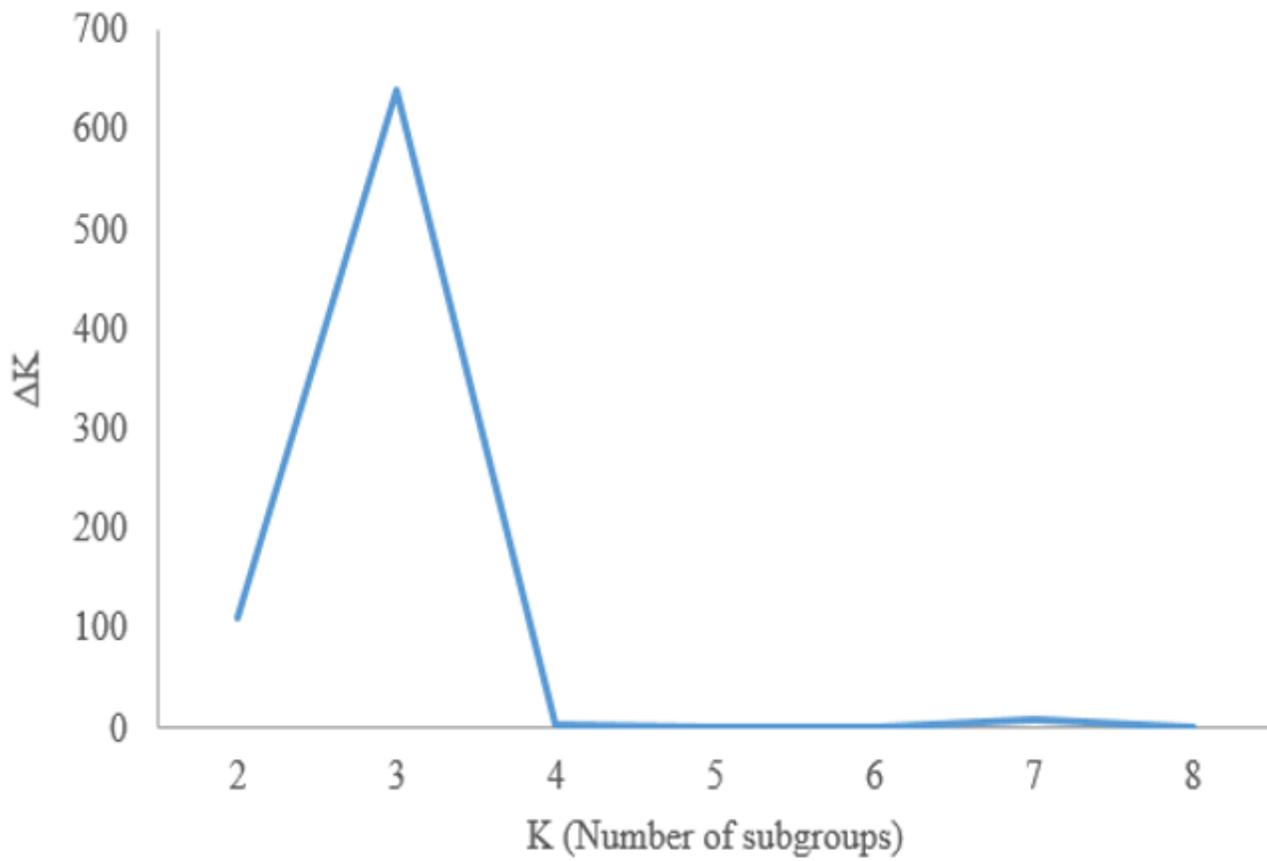


Figure 2

Determination of subpopulations number in wheat genotypes based on ΔK values

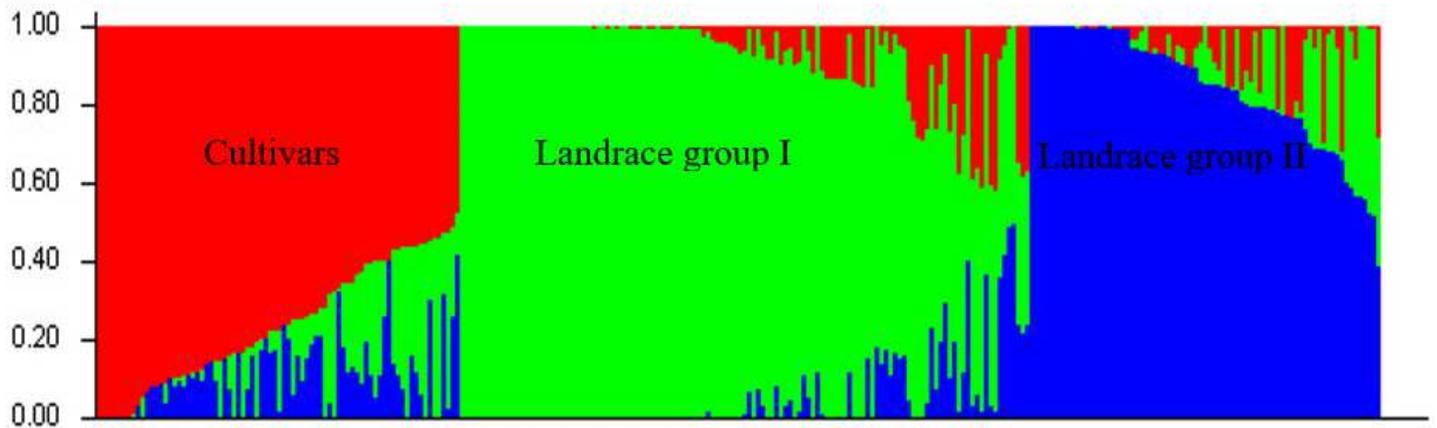


Figure 3

A structure plot of the 282 wheat genotypes and landraces determined by $K = 3$.

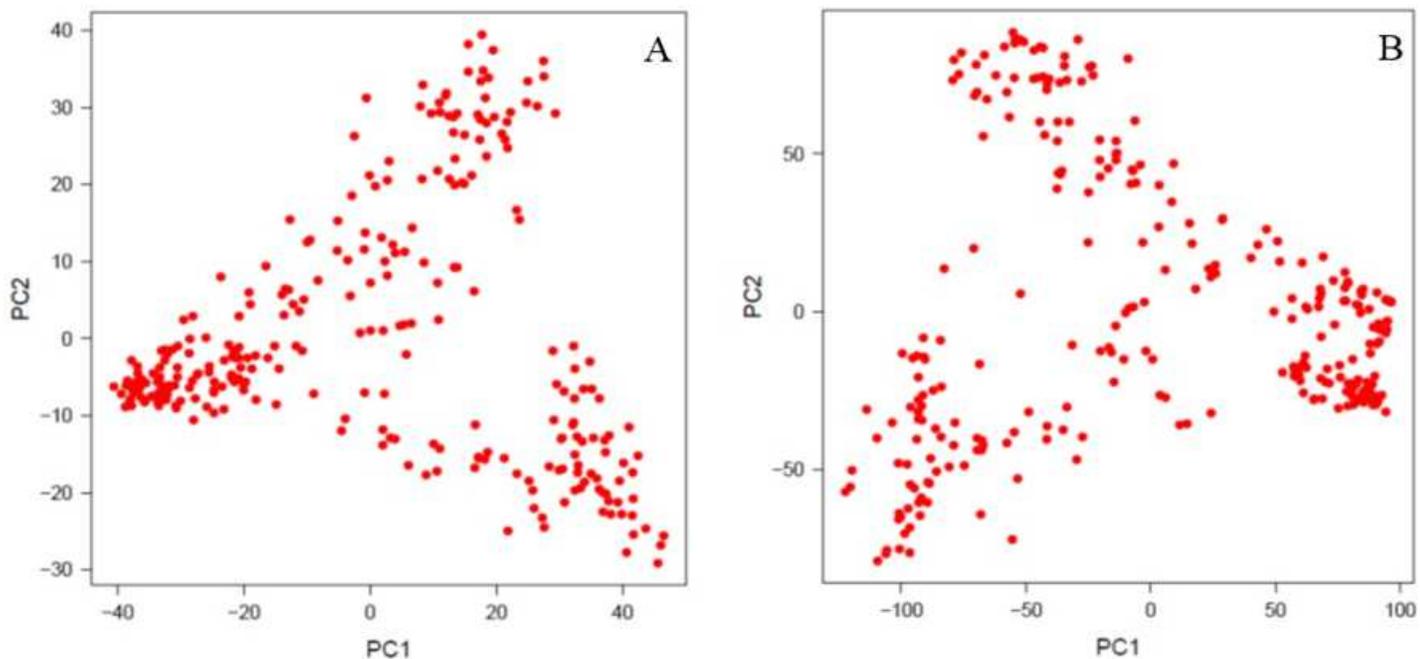


Figure 4

Principal component analysis of the wheat genotypes using original (A) and imputed SNPs (B)

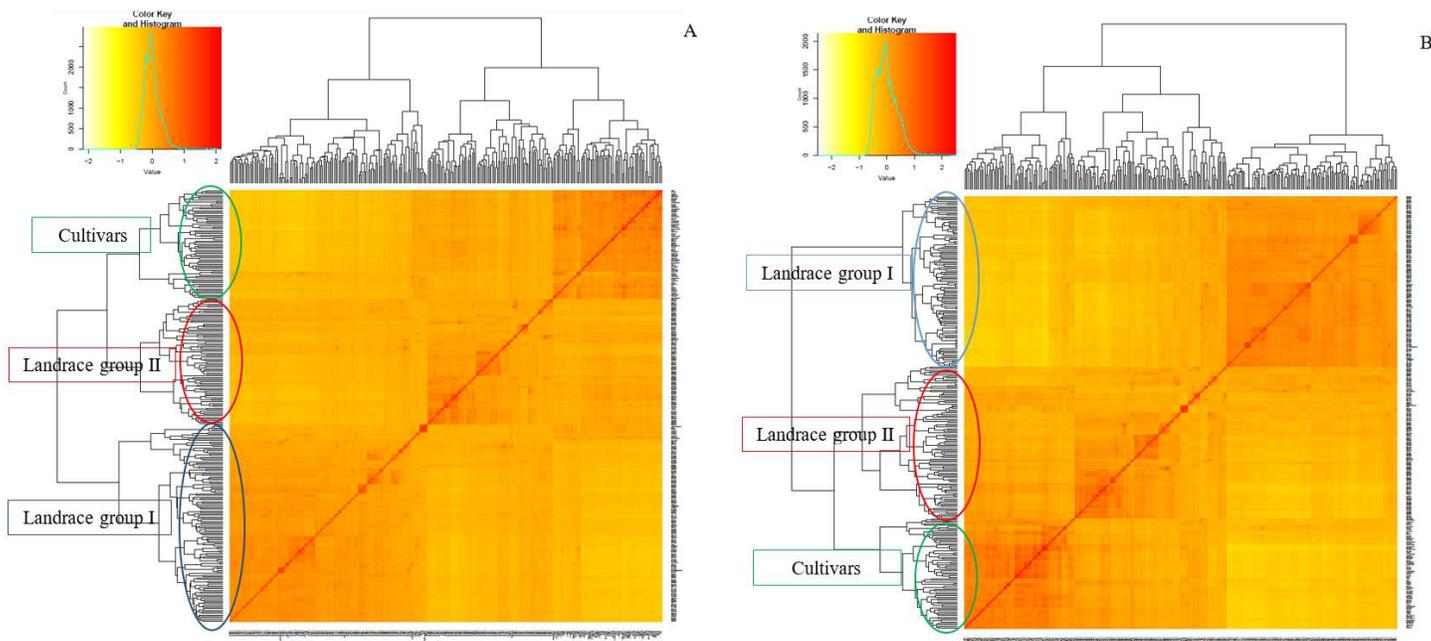


Figure 5

Cluster analysis using Kinship matrix of original (A) and imputed dataset (B) for the wheat genotypes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Genotypeslist.xlsx](#)