Research Article

Association mapping of agronomic traits in bread wheat using a high density 90k SNP array

Muhammad Ali Sher1*, Abdus Salam Khan1*, Zulfiqar Ali1,2 and Sultan Habibullah Khan3

1 Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad
2 Institute of Plant Breeding and Biotechnology, MNS University of Agriculture, Multan
3 Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad
* Correspondence: drabdussalamua@yahoo.com alisherpb@gmail.com

Abstract: In bread wheat the elucidation of genetic loci associated with plant agronomic and yield related traits is an important priority of the breeders. We performed a genome-wide association study (GWAS) using a high-density infinium 90K SNP array. We identified a total of 110 markers traits associations (MTAs) (p ≤0.00) for the following plant traits; days to 50% heading (DH), days to 50% maturity (DM), plant height (Ph) cm, tillers number per plant, spike length (SL) cm and grain yield per plant (GY) g. Most of the SNPs were identified in the A and B genome as compared to the D genome. The significant associated SNPs were mainly distributed on the chromosome 2B, 3B, 6B, and 5B. Nine SNPs on chromosome 5A, 2B and 2D were identified having pleiotropic effects. The Principal Component Analysis was done for six traits in bread wheat genotypes. Two principal components (PC) have more than one eigenvalue and explained 75.98% of the total variation. The population structure analysis differentiated the genotypes into seven subgroups. The correlation analysis showed a significant positive association among SL, NT and GP. Which depicted that these traits are promising for breeding high yielding wheat cultivars. This study provided useful information of the valuable genetic loci for marker-assisted breeding.

Keywords: GWAs; MTAs; SNP; PCA, population structure

1. Introduction

Among cereals crops the bread wheat is the second most important food crop that has direct impact on the daily calories intake by the peoples(www.fao.org/statistics/en/). Wheat crop faces many biotic and abiotic stresses during its growth phases. The unpredicted climate changes make this condition worse. To ensure the food security there is need to breed such wheat genotypes that have high yield potential along with better stress tolerance (1).

Plant yield related traits have complex genetic nature as these are controlled by many genes and these traits are called polygenic traits. Wheat yield has polygenic nature. In wheat crop the genome wide analysis increases the capacity of the scientist for deeper analysis of the polygenic traits. It helps
plant scientist especially breeders for desirable traits pyramiding through the incorporation of the discovered putative genes in the adapted varieties. It is dire need of the time to use modern techniques to fulfill the increasing demand of the human population (2).

The use of conventional/bio-parental mapping population to find the QTLs associated with plant attributes have limitation of low resolution whereas association mapping accounting the linkage disequilibrium (LD) has high resolution power to dissect the genetic basis of polygenic traits as it takes the advantage of diverse genetic background of the natural population. The power, speed and accuracy for identification of the causative allele(s) of different agronomic traits can be enhanced thought the use of genome wide association (GWS). The working principle of genome wide association study is the linkage disequilibrium, the non-random association of different alleles. Plant breeder are frequently using the GWAS for dissecting the genetic basis of the complex traits (3).

In wheat various Marker trait association (MATs) have been conducted, furthermore, around 800 QTLs have been reported over 21 chromosomes in wheat. These QTLs have been identified for plant morpho-physiological and yield related traits (4). Many researchers have identified QTLs using biparental/conventional mapping technique in bread wheat (5) (6),(7-10).

In such studies they have located the caustic loci only limited to the parents used in the cross and thus have low resolution power (11). The identification of genetic loci that are associated with days to 50% heading in bread wheat will enable the researcher to develop early maturing cultivars. The early maturing cultivars have the ability to avoid the terminal heat and drought stress (12).

In wheat breeding the marker assisted selection especially using single nucleotide polymorphism (SNP) has become an important routine tool to improve the cultivars for different biotic and abiotic stress. In recent year a new approach “association mapping” has been developed as a promising technique to identify the genomic regions associated with plant traits. Genome wide association studies (GWAS) has become a choice of method for the identification of genomic regions of the complex quantitative traits in specifically designed plants population or germplasm collections (13).

In GWAS a large panel of population with natural phenotypic variation is used which truly represent the genetic diversity available for different traits in a given region or country (14, 15). One limitation of GWAS is that it does not detect the rare allele (16). However, GWAS is still useful to detect robust QTL that have an effect across different genetic backgrounds and environments (17). This research was initiated with the objective to carry out a genome wide association mapping for different morphological traits and stem reserve carbohydrates mobilization in bread wheat.

In present study we have selected 191 bread wheat genotypes and performed genome wide association mapping using 90K illumine chip.

2. Materials and Methods

2.1. Germplasm collection and sowing
A total of 96 bread wheat genotypes were collected from National Agriculture Research Center, Islamabad. These genotypes were grown according to the randomized complete block design with two replications during the years 2013-2014 and 2014-2015 at the field area of the Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad. Dibbled method of sowing was used and after thinning one plant per hole was retained. The row to row and plant to plant distance was maintained at 30 cm and 15 cm respectively. The plot size was one-meter square for each genotype.

2.2 Phenotypic data collection
At plant maturity the data for the studied traits was recorded in the field conditions from ten randomly selected plants and then averaged. The traits under study were number of days to 50% heading, number of days to 50% maturity, plant height (cm), spike length (cm), tiller number per plant and grain yield (g) per plant.

2.3 Phenotypic data analysis
The estimation of prevalence of genetic variation among the genotypes for the studied traits was done according to (18). Person correlation coefficient was done using the R windows.

2.4 Single-nucleotide polymorphism genotyping
The seed were sown in the pots and after 21 days the young leaf were harvested and DNA was extracted following the protocol given by (19). The 0.8% strength agarose gel was used to check the DNA quality. The NanoDrop 1000 spectrophotometer was used to calculate the DNA concentration. The genotypes for association mapping were genotyped using the Illumina 90K iSelect wheat SNP assay (20).

2.5 Population structure
Population structure was calculated using markers with pairwise $R^2 < 0.5$ for all pairwise comparisons. To assign the subpopulation membership for each genotype, STRUCTURE software version 3.2 was used (21). We used an admixture model with independent allele frequencies, a burn-in of 100,000, and an MCMC replication of 500,000 for $K = 1$–$10$ with five replications. The delta k calculated from the STRUCTURE software was used to select the optimum number of subpopulations (22). AM analysis was conducted using the software TASSEL v.5.0 (23). The mixed linear model (MLM) with PC + Kinship (K) was used for AM, where the genotypic data were filtered for minor allele ($\leq 5\%$) frequency. A total of 14,816 filtered SNPs were used for further AM study. The initial cutoff point for marker–trait association (MTA) was considered at $p \leq 0.001$. Then, this cutoff was subjected to Bonferroni correction (24) to get the threshold ($p \leq 3.4 \times 10^{-6}$). Only the markers identified to be associated in at least two environments were reported.

2.6 Marker–trait association analysis
The mixed linear model (PCA + K) (25, 26) was used for genome wide association mapping through the GAPIT package (11, 22) implemented in R software. The significance association of markers with the traits was declared at $P \leq 0.001$ and the value of $R^2$-marker was used to estimate the magnitude of QTL effects.
3. Results

3.1. Phenotypic data result

The genotypes showed significant genotypic variations (p< 0.05) for all the traits under this study (Table 1). The genotypes × year interaction was found non-significant for all traits. Descriptive statistics data of bread wheat genotypes over the two year 2013-14 and 2014-15 (average data) is presented in Table 2. The mean value for days to 50% heading was 94 and it ranged from 85 to 105 (Table 2). The mean value for days to 50% maturity was recorded 143 whereas the minimum and maximum recorded values for days to maturity were 140 to 155 respectively (Table 2). Plant height of the genotypes ranged 73 to 110.5cm whereas the average calculated value was 95.5cm (Table 2). The minimum and maximum number of tillers per plant were 4 to 9 and the averaged value was 7 (Table 2). The minimum and maximum value of grain yield per plant was ranged 11.5 to 27.0 g and the recorded mean value was 18.5g (Table 2). Correlation coefficient plays a key role in identifying the magnitude and nature of association among different plant agronomic and yield related traits. The correlation results averaged over the two year is depicted in Figure 1. The grain yield per plant was found highly positively associated with number of tillers per plant, spike length and plant height. The NT showed significant positive association with SL and Ph. Days to 50% heading showed positive association with days to 50% maturity.

Table 1. Mean square table for six agronomic traits of 96 wheat genotypes

<table>
<thead>
<tr>
<th>SOV</th>
<th>Df</th>
<th>DH</th>
<th>DM</th>
<th>Ph</th>
<th>NP</th>
<th>SL</th>
<th>GY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>1</td>
<td>106</td>
<td>214.2</td>
<td>115.2</td>
<td>25.9</td>
<td>133.6</td>
<td>222.4</td>
</tr>
<tr>
<td>Genotypes (G)</td>
<td>95</td>
<td>68.4*</td>
<td>157.5**</td>
<td>715.2**</td>
<td>190.1**</td>
<td>1448*</td>
<td>258.5**</td>
</tr>
<tr>
<td>Year (Y)</td>
<td>1</td>
<td>257 n.s</td>
<td>816 n.s</td>
<td>240.2 n.s</td>
<td>119.3 n.s</td>
<td>96.5 n.s</td>
<td>650.3 n.s</td>
</tr>
<tr>
<td>G×Y</td>
<td>95</td>
<td>7.5 n.s</td>
<td>25.4 n.s</td>
<td>70.1 n.s</td>
<td>194.3 n.s</td>
<td>148.2 n.s</td>
<td>190.1 n.s</td>
</tr>
</tbody>
</table>

DH= 50% days to heading, DM= 50% days to maturity, Ph=plant height, NTP= number of tillers per plant, SL= spike length GP = grain yield per plant. * = Significant, ** = Highly significant

Table 2. Descriptive statistics table of the average data 2013-14 and 2014-15 for plant traits

<table>
<thead>
<tr>
<th>Traits</th>
<th>DH</th>
<th>DM</th>
<th>Ph</th>
<th>NT</th>
<th>SL</th>
<th>GY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>94.0</td>
<td>143.0</td>
<td>95.5</td>
<td>7</td>
<td>16.1</td>
<td>18.5</td>
</tr>
<tr>
<td>SE Mean</td>
<td>0.4</td>
<td>0.3</td>
<td>0.7</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>C.V.</td>
<td>11.5</td>
<td>11.8</td>
<td>9.5</td>
<td>7.5</td>
<td>6.1</td>
<td>17.3</td>
</tr>
<tr>
<td>Minimum</td>
<td>85.0</td>
<td>140.0</td>
<td>73.0</td>
<td>4.0</td>
<td>7.0</td>
<td>11.5</td>
</tr>
<tr>
<td>Maximum</td>
<td>105.0</td>
<td>155.0</td>
<td>110.5</td>
<td>9.0</td>
<td>15.0</td>
<td>27.0</td>
</tr>
</tbody>
</table>

DH= 50% days to heading, DM= 50% days to maturity, Ph=plant height, NTP= number of tillers per plant, SL= spike length GP = grain yield per plant.
Figure 1 Pearson’s correlation coefficient for studied traits averaged over the two-year data 2013-14 and 2014-15.

Figure 2 population structure of 96 bread wheat genotypes

3.2. Principal component analysis

The Principal Component Analysis was done for six traits in bread wheat genotypes. Two principal components (PC) have more than one eigenvalue and explained 75.98% of the total variation (Table 3, Figure 3). The PC1 has the highest variability 50.32% followed by PC2 25.55% and PC3 11.66%. As PC1 and PC2 cumulatively explained 75.98% of the total variability, thus the traits falling in these two PC due importance for wheat breeding program. The PC1 was mainly associated with the number of tillers per plant followed by spike length, grain yield per plant and plant height. In PC2 the worth noticing traits were days to 50% maturity and days to 50% heading.

Table 3 Principal component analysis of six traits in 96 wheat genotypes

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>3.02</td>
<td>1.54</td>
<td>0.70</td>
<td>0.32</td>
<td>0.28</td>
<td>0.13</td>
</tr>
<tr>
<td>Variability (%)</td>
<td>50.3</td>
<td>25.6</td>
<td>11.6</td>
<td>5.35</td>
<td>4.67</td>
<td>2.32</td>
</tr>
<tr>
<td>Cumulative %</td>
<td>50.3</td>
<td>75.9</td>
<td>87.6</td>
<td>93.0</td>
<td>97.6</td>
<td>100.</td>
</tr>
</tbody>
</table>
Figure 3 Biplot display of six traits in bread wheat cultivar

3.3. Marker trait association for combined normal and drought stress indices

The MLM approach was used to assess the marker trait association for the following traits under study: DH= days to heading, DM= days to maturity, Ph= plant height, SL= spike length, NT= number of tillers per plant, GY= grain yield per plant.

3.3.1. Days to heading

For days to heading 26 highly significant SNP were found distributed on the chromosomes 1A, 3A, 4A, 6A, 7A, 4B, 5B and 7D. The $R^2$ value for all the significant SNP range from 0.464 to 0.320. Majority of the SNP were distributed on chromosome 1A, 5B and 7A. Among all of the significant SNP the lowest p value was observed for BobWhite_c21482_192 SNP on chromosome 5B (Figure 4).
3.3.2 Days to maturity
For days to maturity a total of 20 highly significant SNP were found distributed on the chromosomes 3A, 4A, 7A, 1B, 2B, 3B, 5B, 6B, 1D, 2D, 3D and 7D. The $R^2$ value for all the significant SNP range from 0.360 to 0.053. Majority of the SNP were distributed on chromosome 1B, 4A and 7D. Among all of the significant SNP the lowest p value was observed for Tdurum_contig46583_2364 SNP on chromosome 4A (Figure 5).

3.3.3 Plant height
For plant height a total of 24 most significant SNP were found distributed on the chromosomes 1A, 4A, 5A, 6A, 2B, 3B, 6B, 1D, 5D and 6D. The $R^2$ value for all the significant SNP range from 0.218 to 0.278. Majority of the SNP were distributed on chromosome 5A, 2B and 6B. Among all of the significant SNP

![Figure 4](image1.png)

**Figure 4** GWAS manhattan plot for days to heading (averaged two year mean data)

![Figure 5](image2.png)

**Figure 5** Genome-wide manhattan plot for days to maturity (averaged two year mean data)
the lowest p value was observed for BS00066569_51 on chromosome 5A (Figure 6).

![Figure 6](image)

**Figure 6** Genome-wide manhattan plot for plant height (averaged two year mean data)

### 3.3.4 Number of tillers/plant

For number of tillers per plant a total of 21 most significant SNP were found distributed on the chromosomes 1A, 5A, 6, 2B, 3B, 7B, 1D 2D and 3D. The R² value for all the significant SNP range from 0.246 to 0.210. Majority of the SNP were distributed on chromosome 2B, 3B and 3D. Among all of the significant SNP the lowest p value was observed for RAC875_c46237_294 SNP on chromosome 3B (Figure 7)

![Figure 7](image)

**Figure 7** Genome-wide manhattan plot for number of tillers/plant (averaged two year mean data)

### 3.3.5 Grain yield per plant

For grain yield per plant a total of 19 most significant SNP were found distributed on the chromosomes 1A, 2A, 3A, 4A, 2B, 3B, 5B, 7B, 2D, 3D and 7D. The R² value for all the significant SNP range from 0.306 to 0.202. Majority
of the SNP were distributed on chromosome 1A, 2B, 3B, 7B and 7D. Among all of the significant SNP the lowest p value was observed for Ku_c4369_351 SNP on chromosome 1A with R^2 value of 0.306 followed by RAC875_c46237_294 SNP on 3B (Figure 8).

**Figure 8** Genome-wide manhattan plot for grain yield per plant (averaged two year mean data)

**Discussion**

It is dire need of the time to break the yield stagnation in the wheat crop through using the modern techniques to fulfill the increasing demands of the food. In this study we have investigated the diverse gene pool of the hexaploid wheat crop and conducted genome wide association mapping for different agronomic and yield related attributes. Some of these traits influence the final grain yield of the plant either directly or indirectly. In association mapping the diversity of the gene pool enhanced resolution power. Genome wide association study mainly detect the association between genetic variability and specific phenotypic variation (11).

The 90 K Illumina iSelect SNP array was used for GWAS analysis after SNP calling for polymorphic. In this research most of the SNPs were identified in the A and B genome as compared to the D genome which showed very few SNPs (27, 28). One reason of this is that the A and B genome have old evolutionary history has compared to the D genome (29-31). The genotypes G39, G32 and G18 were best for grain yield and number of tillers per plant (Figure 3), whereas the genotypes G67, G56 and G25 were poor performer for the GY and NT.

In this study the MTAs identified for days to 50% heading were distributed mainly on chromosome 1A, 5B and 7A. Earlier studies reported similar results like (32) identified MTAs for days to 50% heading on chromosome 2A, 2B, 5B, 7A. (33) identified 5A a major chromosome having highly significantly SNPs associated with DH. Similarly, 15 MTAs on chromosome 5A, 6B, 2B, 2A, 1B and 7D were reported for days to 50% heading in bread wheat under drought tolerance.
Many MTAs were detected for days to maturity on chromosomes 2B, 3B, 4B, 4D, and 6A (33), on chromosomes 2B and 4D in a biparental population (34, 35) identified 114 MTAs for days to maturity that were distributed on 19 chromosomes whereas most of the markers were present on chromosome 4B. In our study majority of the SNP were distributed on chromosome 1B, 4A and 7D.

In our study the mostly significant SNPs were detected on chromosome 5A and 2B for plant height and similar results have been reported by (36-38) used 90K array in Pakistani wheat germplasm and found significant MTAs of the plant height on chromosome 5A.

Earlier studies have detected QTLs for tillering on almost all wheat chromosomes using the traditional QTL mapping approach (39-42) reported a total of 47 significant SNPs associated with number of tillers in bread wheat. (43) reported SNPs on chromosome 2A, 2B, 6A and 7B. (Chen et al., 2017) reported that the markers Excalibur_c40993_1129 on chromosomes 1B, Kukri_c47259_416 and Jagger_c6853_60 on chromosome 2B, RAC875_c14309_317 on chromosome 6B and wsnp_Ex_c14654_22713386 on chromosome 7A were associated with tillering capacity of plant. In our study we detected significant SNPs mainly on chromosome 2B, 3B and 3D.

In cereal crops the final step of any breeding program is grain yield, the polygenic nature of this trait makes it more complex with low heritability and is highly influenced by the environment. (11) reported the most numbers of significant SNPs for yield traits on chromosomes 5A and 6A in a spring wheat population. Similarly, the SNPs detected on chromosomes 5A and 6A are most likely the MTAs reported previously (37, 44) MTA responsible for GY were identified on chromosomes 1AL, 4BL, 6BL, and 7DL, by (38). In our research most of the SNPs were distributed on chromosome 1A, 2B, 3B, 7B and 7D. In marker assisted selection the markers that show pleiotropic affects are quite useful in enhancing the QTLs pyramiding. The marker for plant height GENE-0491_ on chromosome 1A was found associated with number of tillers per plant. Similarly, BobWhite_c31986_324 on chromosome 2B f have pleiotropic effects for plant height and spike length. The Ra_c4096_921 on chromosome 3B for grain yield has pleiotropic effects on NT.

Conclusions:
In this study the identified genetic loci and the MTAs are helpful in pyramiding the desirable traits in bread wheat genotypes. The SNPs that have pleiotropic effect have high value from breeding perspective. A and B genome have high number of SNPs as compared to the D genome.
Endorsement
Dr. Abdus Salam Khan and Dr. Zulfiqar Ali, Department of Plant Breeding and Genetics, University of Agriculture Faisalabad and Dr. Sultan Habibullah Khan, Center for Advanced Studies in Agriculture and Food Security (CAS-AFS), University of Agriculture, Faisalabad (UAF), Pakistan, (CABB-UAF), provided technical assistance and guidance throughout this study.

Interest conflicts
There are no conflicts of interest declared by the authors.

Ethics confirmation
This article does not include any experiments with humans or animals conducted by any authors.

References


