High-Throughput Digital Imaging Analysis Of Historical Wheat Cultivars From Pakistan

Tehreem Tahir  
Mohi-ud-Din Islamic University

Awais Rasheed  
Quaid-i-Azam University Islamabad: Quaid-i-Azam University

Sadaf Kayani  
Mohi-ud-Din Islamic University

Asim Shahzad (agr12@yahoo.com)  
Mohi-ud-Din Islamic University  
https://orcid.org/0000-0003-2740-7233

Research Article

Keywords: Digital Imaging, Principal Component Analysis (PCA)

Posted Date: July 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3123747/v1

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

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on December 11th, 2023. See the published version at https://doi.org/10.1007/s10722-023-01801-7.
Abstract

Wheat is the most significant cultivated crop in Pakistan, ranking first in terms of acreage and production among all cereals. Wheat grain morphology has been harvested and manipulated since the dawn of agriculture, and it remains a major breeding focus today. To establish, genetic basis of phenotypic variation in wheat grain morphology, the current research was conducted to develop a phenotyping method based on digital imaging to capture the grain morphology in bread wheat cultivars of Pakistan and to characterize variation in grain morphology in Pakistani wheat cultivars and their association with TaCW genes. In this research 64 wheat cultivars were analysed by digital imaging approach. The images were captured using a flatbed scanner with a transparency adapter, and the imaging software image J 1.27 was used for analysis. The cultivars used in this study showed larger variation in grain morphology in more than fifteen descriptors. Measured dimensions included grain area, perimeters, height, width, circularity and solidity. Minimum grain area was showed by the T9 and maximum grain area was shown by the Pirsabak-2013, minimum grain perimeter was showed by the T9 and maximum grain perimeter was showed by the Barsat. Similarly, maximum grain area was exhibited by the Markaz-2019 and minimum grain area was shown by the C-271. Maximum grain perimeter was showed by the PARI-73 and minimum grain perimeter was showed by the C-271 respectively, under well-watered (WW) conditions. The coefficient of determination was also higher indicating the reliability and consistency of digital imaging method. Digital imaging allowed many aspects of development, functions, traits and health to be monitored and traced in ways previously unattainable. Our results are encouraging that large image data sets of grains can be processed by high throughput and accurate manner.

1. Introduction

The most widely grown cereal crop is wheat (*Triticum aestivum*). It is the most important cereal for white race so far in temperate region. This is an annual herb, of a genus called *Triticum*, which includes a considerable number of both wild and cultivated plants. It grows 215 million hectares a year (Li et al 2020: FAO, 2011). It is the primary staple food for 2.5 billion people and the first of three big crops of cereals. In china yield increases 140-160kg per hectare can be achieved when TKW increases by a mere 1 gram (Zhang et al 2022). Global wheat production was 716 million tons in 2013, which is higher than in 2007 (Acharya et al 2017: CYMMYT, 2013). The Use of agriculture practices not only increases production but also reduce the risk posed by environmental and economic factors (Chauhan, 2012). Improving wheat productivity is a biggest challenge for the world. The improved yield is resistant to biotic stress including drought, high temperature and salinity etc. wheat is a carbohydrate source for the majority of the country and the world's leading dietary protein source in human food, with a protein content of around 13 percent which is comparatively high when associated to additional main cereal crops.

Wheat (*Triticum aestivum L.*) breeding in Pakistan began in the early 1930s, before partitioning in the United India, and was accelerated after the green revolution with over 68 cultivars published so far (Ahmad et al 2005). The increase in production is primarily due to encroachment on other crops with hardly any increase in productivity per acre due to an increase in the minimum support price. Fortunately, farmers are also unhappy with higher cost of output as a result of duties on agricultural inputs, the energy torisis and a slump in the value of the currency. It is also true that the productivity of wheat at present is not more than 2893mt/ha whereas in order to feed the increasing population at an annual rate of 1.8% and to sustain the feed and seed requirements at 10% of the total annual output and 1 million tons of strategic reserves, the level of production of wheat must rise to 28.2 million tons by 2020 by raising the productivity of integrated approach (Vu et al 2020). In Pakistan, 152 wheat varieties have been published to date, with a large number in the last decade, with a share of Punjab (69), Sindh (25), KP (44) and Baluchistan (8). During this time, only a few varieties become popular as mega varieties. Most varieties were depleted within 3–5 years of release due to rust susceptibility or other factors, and their names can be found in the articles. In Pakistan, more than 13 wheat breeding programmes are active (Zahoor et al 2017)

As the world's population grows, food security is increasingly dependent in our ability to improve grain yield through crop breeding (Gupta et al., 2015). The TGW an important agronomic trait, and is significantly related to yield. The trait is primarily determined by grain weight (GW), grain length (GL), and grain thickness (GT), all of which are positively correlated (Hu et al., 2016). As a result, increasing grain weight is a constant goal of wheat breeding programmes. Grain size determines grain weight and has an impact on grain quality. In wheat, the size of grain is an essential component of grain production, and the large size of grain is a key feature of domestication and procreation (Ogutcen et al 2018). New grain varieties have greater grain width and lower grain longitude relative to old grain varieties which display greater variation in granulated feed (Du et al 2021) therefore genes or QTL related to grain size and forms that are of interest of domestication and procreation purposes. The weight of mille grains depends on grain weight, length and thickness of grain(Cabraal et al 201) and all of this represent thousand grain weight, so size of grain improves the weight of grain, in wheat breeding programs.

Grain morphology are vital characteristics that influence the market price of wheat grain because they affect the efficiency of the grinding process i-e flour quality and yield (Hassoon et al., 2021), according to theoretical models, optimising grain morphology with large and round grains would increase milling yield. Many other industry-used consistency parameters are influenced by grain morphology. Before milling, specific weight (kg of weight per litter of bulk grain) is usually used to grade wheat and is assumed to be related to the shape or size of the grain because these parameters determine grain packs are milled. Many other consistency parameters used by the industry are influenced by grain morphology (Ponce-Garcia et al 2017). The genetic and phenotypic variations in wheat grain morphology, while genetically and developmentally important, are astonishingly understudied because this trait is difficult to calculate. Previous studies used a limited number of measurements, which were individually examined in population maps (Hasson et al 2010 and Kumari et al 2018). One strategy is to combine the dissimilar metrics into a low-dimensional structure (an uncommon variables that capture most of the variance of the trait) and then use it to describe the genomic basis of the phenotypic relationship between grain size and shape. To develop a better understanding of the inherited basis of grain size and shape variation, researchers used numerous different recombinant populations found in the wheat germplasm pool (Haliloglu et al 2023). In addition to the phenotypic structure of the traits and the degree of variation retained in wheat domestication, grain material from accessions of primitive wheat species and modern wheat varieties was evaluated we show grain morphology is largely independent of grain types in both DH population and primitive wheat species, and that there is a substantial reduction in phenotypic variation in grain shape in the breeding pool, which could be attributed to the relatively bottleneck. This phenotypic structure is the result of a unique genetic architecture in which similar hereditary components are involved in the regulation of specific characteristics in dissimilar wheat varieties (Bhati et al., 2022).
With an increasing negative impact of climate change, ensuring global food security through sustainable wheat production for the projected population is a major challenge (Farooq et al., 2023). It is therefore critical to increase production continuously, primarily through higher yields. Wheat breeding programmes are concentrating on the development of high yielding varieties with good end-use quality. A number of genes interact with each other and the environment to control grain yield (GY) (Mahjourimajd et al., 2016). Due to the use of Rht genes in wheat breeding, grain yield has increased over the last four decades due to increased grains per square meter or larger grain sizes (Feng, et al 2021). Wheat yield increases have slowed significantly in recent years due to lack of breakthrough germplasms and breeding methodologies (Bapela et al., 2022). Grain yield has been shown to be a dynamic and quantitative trait regulated by a number of low heritability genes and to have a significant environment impact, making breeding programmes difficult to exploit and develop (Laugerotte et al., 2022). The number of spikes per m², the number of grain per spike, the number of grains per m² and the weight of 1000 kemes are all used to calculate grain yield in bread wheat. For the source of variation in both genetic and environmental factors, significant negative associations between these wheat varieties have already been established. In general, the difference between (a) the numbers of spikes per m² and (b) the number of grain per m². Just before reproductive phase, the number of grains per m² (the product of the number of spikes per m² and the number of grain per meter spike) is calculated which shown to be the most vital factor in grain harvest, especially under ideal conditions (Beral, Renaud, and Vincent, 2020).

The grain quality of wheat is essential to the welfare of humans, in contrast to yield. Wheat is commonly pragmatic in many foods and industrial uses because of its high protein content (GPC). Grain quality currently at the forefront in crop sciences, and improving wheat grain protein content and processing efficiency has become one of the main breeding goals (Chen et al., 2022). Wheat end-use quality is determined by grain hardness or endosperm texture, protein content, and protein quality. Higher grain quality requirements are increasing exponentially, as an outcome of innovative production methods, dramatic demographic shifts. Breeders must use non-traditional quality management methodologies because there is a strong selection pressure on favourable quality alleles. The introduction of a number of disciplines, such as genomics, biotechnology, and genetic resource utilisation has sparked interest in determining the genetic, biotechnological, and physiological basis of quality encoding characteristics in wheat (Rasheed et al., 2014).

Size of wheat grains are the essential characteristics, determined by genetics and environmental factors that cause variations in those attributes. Various method of classification use different morphological attributes for the literature documented the sorting of various cereals grains and varieties through image processing technique. There are many ways which can determine seed size and shape (e.g.) it can be measure by callipers or computationally unless the very small size can be measured (Tanabata et al 2012) but Digital Image Analysis (DIA) is the methodology of physical image transformation into digital values (Riber-Hansen et al 2012). The exact size of the grain was determined was by the (Rasheed, 2014) usage of Image J software kit. (JAMIL et al 2017) Image J is also used to regulate the magnitude of the grain. Digital Imaging technique can be used to highlight even minor seed changes. The DIA can be based on the pixel quantity. They can lead to a change in the wide range of measureable data and have cast-off in many previous works (Yang et al 2021). The measurement of seed shape characteristics is less accurate and laborious without digital assistance. Tri-dimensional wheat seed image data cannot be obtained without high throughput digital imaging. DIA can be used to calculate the size and shape reduction in any dimension. Prior to their use, new variations in untapped germplasm require state-of-the-art phenotypic characterization. High-throughput genotyping does not align phenotypic development. To get the most out of low-cost genotyping tools, more precise, efficient and high throughput phenotyping pathways must be investigated (Houle et al., 2010). The use of DIA for high-throughput wheat grain size and shape phenotyping has recently been demonised (Williams's et al., 2012).

Wheat grain morphology is important in increasing wheat yield, but there aren't many platforms for precisely determining grain size and weight have been recognised, but their presence in Pakistani wheat cultivars is unknown. Keeping in view the importance of seed the present study was designed to optimize a phenotyping method based on digital imaging to capture the grain morphology in bread wheat cultivars of Pakistan and to characterize variations in grain morphology in Pakistani wheat cultivars and their association with TaCWI.

Results

3.1 High Throughput Digital Image Analysis:

The 64 Historical Wheat Cultivars were analysed by the digital imaging approach. The Basler machine vision sensor provided higher resolution and ready to analyse wheat seeds. In total, 64 images were captured and analysed. All the images were analysed within 60 seconds. Each descriptor were analysed at significance of P < 0.001. First, pixels per mm was calculated by the image J software by placing a known scale inside the image. In this case, all the images in this platform had 144 distance in pixels or 144 pixels per cm with 1.00 known distance. Set measurements by selection (Area, Standard deviation, minimum values, maximum, limit threshold, Perimeters, Ferrets, and display label). The output file was mask file and mask file had been cleaned for any anomalies. Some analysed measurements are taken which are then exported to MS excel for further analysis of the data. Further Tables 2 and 3 shows basic statistics of grain morphology. The summary statistics of grain size and shape phenotypes under (WL) and (WW) conditions are; min (minimum value), max (Maximum value), SE.mean (Standard error), Var. (Variance), Std.dev. (Standard deviation) and Coef. var. % (percentage Coefficient of variation). The summary statistics is shown in Table 2 (a) and 2 (b). The Table 2 (a) shows statistics of grain size and shape under water limited conditions which are as follows; the average grain area under water limited conditions was 0.20 with a range of 0.15 and 0.24cm². Minimum grain area was shown by T9 and maximum grain area was shown by Pirsabak-2013. The average grain perimeter under water limited condition was 2.07 with a range of 1.71 and 2.35cm. Maximum grain perimeter was showed by the Barsat which was 2.35 and the minimum grain perimeter was showed by the T9 which was 1.71. The average grain height under water limited condition was 0.57 with a range of 0.49 and 0.63cm. Maximum height was calculated for Chakwal-86 and minimum was recorded for Wafaq-2000. The average grain solidity was 0.92 with a range of 0.90 and 0.94cm². Maximum grain solidity was recorded for T9 which was 0.90 and minimum grain solidity was recorded for Zincol-2016 and Barsat. The average grain circularity under water limited condition was 0.65 with a range of 0.59 and 0.740cm². Maximum grain circularity was showed by the Wafaq-2004 and minimum grain circularity was Barsat. The Table 2 (b) shows statistics of grain size and shape under well-watered conditions which are as follows; the average grain area under well-watered conditions was 0.20 with a range of 0.15 and 0.24cm². Maximum grain area was shown by Markaz-2019 and minimum grain area was shown by C271 under well-watered conditions. The average grain perimeter was 2.07 with a range of 1.71 and 2.35cm². Maximum grain perimeter was recorded for PARI-73 which was 2.54 and minimum was shown by C271 which
was 1.692. The maximum and minimum grain height, circularity and solidity was recorded in same variety which was PARI-73 for minimum and C-271 for maximum. The average grain angle under well-watered conditions was 91.9 with a range of 81.5 and 102.9 cm. Minimum grain area was shown by NARC-2009 which was 79.5 and maximum was shown by Punjab-2011 which was 99.3. The box plots are shown for all traits in Fig. 4. All the phenotypes were higher for well-watered conditions (WW) except circularity, solidity which were higher in water limited (WL) conditions. The box plot showed that there were a greater number of outliers in X, Y, XM, YM, BX and BY as compared to other phenotypes. The histogram for each trait are shown in Fig. 3 (b). The histogram showed the normal distribution for all phenotypes but higher frequency recorded in water limited (WL) conditions as compared to well-watered (WW) conditions.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.15</td>
<td>5.95</td>
<td>5.20</td>
<td>5.96</td>
<td>5.20</td>
<td>1.71</td>
<td>5.67</td>
<td>4.90</td>
<td>0.47</td>
<td>0.49</td>
<td>0.62</td>
<td>81.50</td>
<td>0.59</td>
<td>0.64</td>
<td>39.78</td>
</tr>
<tr>
<td>Max</td>
<td>0.24</td>
<td>15.3</td>
<td>11.1</td>
<td>15.3</td>
<td>11.1</td>
<td>2.35</td>
<td>15.0</td>
<td>10.89</td>
<td>0.63</td>
<td>0.63</td>
<td>0.76</td>
<td>102.9</td>
<td>0.70</td>
<td>0.81</td>
<td>62.61</td>
</tr>
<tr>
<td>Range</td>
<td>0.08</td>
<td>9.33</td>
<td>5.95</td>
<td>9.33</td>
<td>5.95</td>
<td>0.64</td>
<td>9.33</td>
<td>5.98</td>
<td>0.15</td>
<td>0.14</td>
<td>0.14</td>
<td>92.17</td>
<td>0.70</td>
<td>0.73</td>
<td>51.86</td>
</tr>
<tr>
<td>Median</td>
<td>0.20</td>
<td>11.4</td>
<td>9.20</td>
<td>11.4</td>
<td>9.20</td>
<td>2.06</td>
<td>11.1</td>
<td>8.92</td>
<td>0.56</td>
<td>0.57</td>
<td>0.35</td>
<td>90.81</td>
<td>0.64</td>
<td>0.73</td>
<td>52.10</td>
</tr>
<tr>
<td>Mean</td>
<td>0.20</td>
<td>11.4</td>
<td>9.20</td>
<td>11.4</td>
<td>9.20</td>
<td>2.06</td>
<td>11.1</td>
<td>8.92</td>
<td>0.56</td>
<td>0.57</td>
<td>0.69</td>
<td>91.17</td>
<td>0.65</td>
<td>0.73</td>
<td>52.10</td>
</tr>
<tr>
<td>SE.mean</td>
<td>0.00</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
<td>0.02</td>
<td>0.29</td>
<td>0.21</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>22.24</td>
<td>0.00</td>
<td>0.00</td>
<td>23.59</td>
</tr>
<tr>
<td>Var</td>
<td>0.00</td>
<td>4.26</td>
<td>2.12</td>
<td>4.26</td>
<td>2.12</td>
<td>0.02</td>
<td>4.25</td>
<td>2.14</td>
<td>0.09</td>
<td>0.01</td>
<td>0.01</td>
<td>22.24</td>
<td>0.00</td>
<td>0.00</td>
<td>23.59</td>
</tr>
<tr>
<td>Std.dev</td>
<td>0.01</td>
<td>2.06</td>
<td>1.45</td>
<td>2.06</td>
<td>1.45</td>
<td>0.14</td>
<td>2.06</td>
<td>1.46</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>4.716</td>
<td>0.02</td>
<td>0.03</td>
<td>4.857</td>
</tr>
<tr>
<td>Coef.var</td>
<td>9.3%</td>
<td>18%</td>
<td>16%</td>
<td>18%</td>
<td>16%</td>
<td>7%</td>
<td>18%</td>
<td>16%</td>
<td>7%</td>
<td>5.4%</td>
<td>6%</td>
<td>4%</td>
<td>5.2%</td>
<td>5.1%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

### Table 3

Allelic effect of the TaCwi-A1 gene on various grain morphology parameters in Pakistani historical wheat cultivars.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
<th>Area (cm)</th>
<th>Grain Length (cm)</th>
<th>Grain Width (cm)</th>
<th>Roundness</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaCwi-A1a</td>
<td>31</td>
<td>0.22a</td>
<td>0.73a</td>
<td>0.36a</td>
<td>0.502ns</td>
</tr>
<tr>
<td>TaCwi-A1b</td>
<td>29</td>
<td>0.19b</td>
<td>0.68b</td>
<td>0.34b</td>
<td>0.503ns</td>
</tr>
</tbody>
</table>

Students t-test Significance at P < 0.05

### 3.2 Coefficient of Correlation among grain traits

The coefficient of correlation was determined for both traits. The data are shown in Fig. 4. There was an expressively higher correlation between grain area and width (r = 0.86), grain height (r = 0.85), whereas there was a negative correlation between grain area and circularity (r = -0.56). Similarly, there was higher correlation between grain perimeter and width (r = 0.88), grain height (r = 0.87), grain ferret (r = 0.91) while there was negative correlation between grain perimeter and solidity (r = 0.74) and also negatively correlated with circularity (r = -0.11). There was higher correlation between grain weight and grain height (r
The seed detection method, the degree of overlap between seeds, and the intensity of removing awns and pedicels all have parameters set stored in the designated folder. It can detect overlapping seeds automatically and delete awns and pedicels automatically, which are removed from the

Since there was a positive correlation between variables, the direction of all variables was on each side. Some cultivars were closed to a variable showing importance for that specific trait and some cultivars was not close to variables. The scree plot showing variation explained by each variable is shown as Fig. 5(c) and (d). Since, there was positive correlation between all variables, the direction of all variables was on all sides. Some cultivars was closed to a variable showing importance for that specific trait. For example, Pak-81, Shahkar and inqilab-91 was close to round variable whereas Rawal-87, Atta habib, C-271, Chakwal-86 were away from round variable but showing importance towards the specific trait. WL-711 was close to width and ferret variable. Pari-73 was close to minor variable. NARC-2009 was close to area. Zincol-2016 was close to minor axis and also close to area. Pakistan-2013 was close to area whereas C-273, Barsat, Galaxy-2013, Punjab-96 and Markaz-2019 was not close to height, Feret and perimeter but cultivars are showing importance toward these variables.

The student t-test was performed at a significance of \( P < 0.05 \), as shown, Table 3. The allelic frequency is higher in TaCwi-A1a whereas lower in TaCwi-A1b and are further discussed below. (GL) grain length (GA) grain area (GW) grain weight. Table 3 and Fig. 6 shows as; The wheat CWI-21 and CWI-22 primers were used to identify the allelic states at wheat cell wall invertase (TaCwi-A1a) gene in historical wheat cultivars of Pakistan. The frequency of TaCwi-A1a favoured grain size traits and present in 31 varieties, while TaCwi-A1b was unfavourable for grain size traits and was present in 29 varieties. The allelic effect was significant for grain length, width, area and perimeter, while there was non-significant effect on grain roundness.

### 3.3 Multivariate Analysis Using Principal Component Analysis (PCA)

The variability of grain attributes could be quantified using principal component analysis. The PCA decomposes variances in morphology of grain into mutually independent quantitative features, which are the principal components. The PCA biplot analysis was used to identify the most appropriate combination of the characteristics considered for grains, as shown in Figs. 5 (a) and 5 (b). Showing variables and vectors and cultivars as the point of the PCA biplot. The first principal component under water limited condition (WL) accounted for 43.7% of the total variation followed by 24.4% of the total variation. Since there was a positive correlation between variables, the direction of all variables was on each side. Some cultivars were closed to a variable showing importance for that specific trait and some cultivars was not close to variables. The scree plot showing variation explained by each variable is shown as Fig. 5 (c) and (d). Since, there was positive correlation between all variables, the direction of all variables was on all sides. Some cultivars was closed to a variable showing importance for that specific trait. For example, Pak-81, Shahkar and inqilab-91 was close to round variable whereas Rawal-87, Atta habib, C-271, Chakwal-86 were away from round variable but showing importance towards the specific trait. WL-711 was close to width and ferret variable. Pari-73 was close to minor variable. NARC-2009 was close to area. Zincol-2016 was close to minor axis and also close to area. Pakistan-2013 was close to area whereas C-273, Barsat, Galaxy-2013, Punjab-96 and Markaz-2019 was not close to height, Feret and perimeter but cultivars are showing importance toward these variables. Pirsabak-2004 was close solidity, Ahsan-2016 was close to circularity, Bathor and Maxipak was close to solidity whereas T9 and Wafaq-2000 was away from solidity but still these cultivars having solidity variables. The Fig. 5 (b) shows that under well-watered conditions (WW), 41.8 percent of the total variations were accounted for, followed by 25.5 percent of the total variation. The direction of variables was on three sides except one side. Some cultivars was closed to a variables and some are away from variables. For example, Inqilab-91 was close to round variable, NARC-2009 was close to solidity, NARC-2011 was close to solidity variable, Pothowar-700 was close to circularity variable, Parwaz-94 was close to angle variable, Barani-83 was close to minor variable, Suleman-96 was close to perimeter variable, Bakhar-83 was close to angle variables, AS-2002 was close to solidity variable, GA-2002 was close to solidity variable, Chakwal-50 was close to X, GA-2002 was close to height variable and PARI-73 was close to minor variable. C-271 was away from variables such as circularity. Markaz-2019, Auqab-2002 and Gold-2016 are away from variables but still fall in a group of ferret, width and major variables, this shows the importance of variables toward the cultivars. Miraj-2008, MH-97 and Millet-2011 was not close to any of variables. The scree plot of the main components of the analysis in multivariate statistics. The scree plot is used to determine how many factors to keep in the principal components for the principal component analysis. Figures 5(c) and 5(d) depict a scree plot illustrating variation explained by each variable under WL and WW conditions.

### 3.4 Allelic Effect of TaCwi gene on grain morphology

The student t-test was performed at a significance of \( P < 0.05 \), as shown, Table 3. The allelic frequency is higher in TaCwi-A1a whereas lower in TaCwi-A1b and are further discussed below. (GL) grain length (GA) grain area (GW) grain weight. Table 3 and Fig. 6 shows as; The wheat CWI-21 and CWI-22 primers were used to identify the allelic states at wheat cell wall invertase (TaCwi-A1a) gene in historical wheat cultivars of Pakistan. The frequency of TaCwi-A1a favoured grain size traits and present in 31 varieties, while TaCwi-A1b was unfavourable for grain size traits and was present in 29 varieties. The allelic effect was significant for grain length, width, area and perimeter, while there was non-significant effect on grain roundness.

### 2. Materials and methods

#### 2.1 Sample collection

Plant material consists of 64 Historical wheat cultivars which have been released in Pakistan since 1965. The wheat cultivars are acquired from the National Agriculture Research Centre, Islamabad. Seeds are used from the fresh harvest of the field trial conducted during the wheat growing season of 2019-2020. The detailed name of cultivars are shown in Table 1.

#### 2.2 Phenotyping:

Various wheat cultivars where used for imaging study, using the procedure, discussed in (William et al., 2013). Wheat grains from each cultivars placed horizontally positioned at the appreciate distances to the LED backlight. Quality digital images ~60 pixels/mm have been taken. All photos are named according to the genotype accession number.

#### 3.2.1 Imaging Software

All images will be cropped using Infra View software (www.ifraview.com) to only have the kemels and the normal size. Contrast and brightness of the image will be improved to reduce shadowing errors in edge detection. The entire editing was done by batch converter which is computer command. All photos are stored in the designated folder. It can detect overlapping seeds automatically and delete awns and pedicel automatically, which are removed from the analysis. The seed detection method, the degree of overlap between seeds, and the intensity of removing awns and pedicels all have parameters set.

#### 2.3. Image J software:
The image J software from the national Institutes of Health NIH in the United States was used to measure images, directly from JPEG files, that were opened in the Image J, set to 8-bits and adjusted to the colour threshold to avoid measuring false positives. The quantitative measures will be derived from adjusted images, and with each photograph Image J output a global scale using the normal size will be built-in for measurement of image sets to be transferred to Microsoft Excel 365 for data analysis.

2.4 Data analysis:

Each descriptor was analysed for significance at P=0.001 using the F-test. PROC GLM procedure in SAS version 9.0 procedure used for the significance test and the co-efficient of correlation. Student T-test used to find out whether TaCWI alleles are correlated with grain morphology traits.

3. Results

3.1 High Throughput Digital Image Analysis:

The 64 Historical Wheat Cultivars were analysed by the digital imaging approach. The Basler machine vision sensor provided higher resolution and ready to analyse wheat seeds. In total, 64 images were captured and analysed. All the images were analysed within 60 seconds. Each descriptor were analysed at significance of P<0.001. First, pixels per mm was calculated by the image J software by placing a known scale inside the image. In this case, all the images in this platform had 144 distance in pixels or 144 pixels per cm with 1.00 known distance. Set measurements by selection (Area, Standard deviation, minimum values, maximum, limit threshold, Perimeters, Ferretes, and display label). The output file was mask file and mask file had been cleaned for any anomalies. Some analysed measurements are taken which are then exported to MS excel for further analysis of the data. Further Table 2 and 3 shows basic statistics of grain morphology. The summary statistics of grain size and shape phenotypes under (WL) and (WW) conditions are; min (minimum value), max (Maximum values), SE.mean (Standard error), Var. (Variance), Std.dev. (Standard deviation) and Coef. var. % (percentage Coefficient of variation). The summary statistics is shown in Table 2 (a) and 2 (b). The Table 2 (a) shows statistics of grain size and shape under water limited conditions which as follows; the average grain area under water limited conditions was 0.20 with a range of 0.15 and 0.24cm². Minimum grain area was shown by T9 and maximum grain area was shown by Pir sabot-2013. The average grain perimeter under water limited condition was 2.07 with a range of 1.71 and 2.35cm². Maximum grain perimeter was showed by the Barsat which was 2.35 and the minimum grain perimeter was showed by the T9 which was 1.71. The average grain height under water limited condition was 0.57 with a range of 0.49 and 0.63cm². Maximum height was calculated for Chakwal-86 and minimum was recorded for Wafaq-2000. The average grain solidity was 0.92 with a range of 0.90 and 0.94cm². Maximum grain solidity was recorded for T9 which was 0.90 and minimum grain solidity was recorded for Zincol-2016 and Barsat. The average grain circularity under water limited condition was 0.65 with a range of 0.59 and 0.740cm². Maximum grain circularity was showed by the Wafaq-2004 and minimum grain circularity was Barsat. The table 2 (b) shows statistics of grain size and shape under well-watered conditions which are as follows; the average grain area under well-watered conditions was 0.20 with a range of 0.15 and 0.24cm². Maximum grain area was shown by Markaz-2019 and minimum grain area was shown by C-271 under well-watered conditions. The average grain perimeter was 2.07 with a range of 1.71 and 2.35cm². Maximum grain perimeter was recorded for PARI-73 which was 2.54 and minimum was shown by C271 which was 1.692. The maximum and minimum grain height, circularity and solidity was recorded in same variety which was PARI-73 for minimum and C-271 for maximum. The average grain angle under well-watered conditions was 91.9 with a range of 81.5 and 102.9cm². Minimum grain area was shown by NARC-2009 which was 79.5 and maximum was shown by Punjab-2011 which was 99.3. The box plots are shown for all traits in Figure 3 (a). All the phenotypes were higher for well-watered conditions (WW) except circularity, solidity which were higher in water limited (WL) conditions. The box plot showed that there were a greater number of outliers in X, Y, XM, YM, BX and BY as compared to other phenotypes. The histogram for each trait are shown in Figure 3 (b). The histogram showed the normal distribution for all phenotypes but higher frequency recorded in water limited (WL) conditions as compared to well-watered (WW) conditions.

3.2 Coefficient of Correlation among grain traits

The coefficient of correlation was determined for both traits. The data are shown in Figure 4, There was an expressively higher correlation between grain area and width (r = 0.86), grain height (r = 0.85), whereas there was a negative correlation between grain area and circularity (r = -0.56). Similarly, there was higher correlation between grain perimeter and width (r = 0.88), grain height (r = 0.87), grain ferret (r = 0.91) while there was negative correlation between grain perimeter and solidity (r = 0.74) and also negatively correlated with circularity (r = -0.11). There was higher correlation between grain weight and grain height (r = 0.67), highly correlated with minor axis (r = 0.75) but negatively associated with angle (r = -0.11) circularity(r = -0.59), round (r = -0.15) and solidity (r = 0.54). The grain width is positively associated with ferret, minor axis, major axis and height. Thus, height is positively associated with major, minor, angle and ferret but negatively associated with circularity, round and solidity. The ferret only has negatively correlated with roundness and solidity. Grain roundness has a clear negative correlation with grain height, implying that the two traits affect grain weight in different ways. Parallel and perpendicular deviations from the optical ellipse were also positively correlated with grain length and width, implying the unconventionality from the ellipse increases grain length and width. The length and width of the grains had a small positive relationship, indicating that certain cultivars had larger and longer grains. Apart from perfect correlation in derived variables, the higher correlation was found between Ferret and major axis (r=0.98) which was positively correlated.

3.3 Multivariate Analysis Using Principal Component Analysis (PCA)

The variability of grain attributes could be quantified using principal component analysis. The PCA decomposes variances in morphology of grain into mutually independent quantitative features, which are the principal components. The PCA biplot analysis was used to identify the most appropriate combination of the characteristics considered for grains, as shown in Figures 5 (a) and 5 (b). Showing variables and vectors and cultivars as the point of the PCA biplot. The first principal component under water limited condition (WL) accounted for 43.7% of the total variation followed by 24.4% of the total variation. Since there was a positive correlation between variables, the direction of all variables was on each side. Some cultivars were closed to a variable
showing importance for that specific trait and some cultivars was not close to variables. The scree plot showing variation explained by each variable is shown as Figure 5 (c) and (d). Since, there was positive correlation between all variables, the direction of all variables was on all sides. Some cultivars was closed to a variable showing importance for that specific trait. For example, Pak-81, Shahkar and inqilab-91 was close to round variable whereas Rawal-87,Atta habib, C-271, Chakwal-86 were away from round variable but showing importance towards the specific trait. WL-771 was close to width and ferret variable. Pari-73 was close to minor variable. NARC-2009 was close to area. Zincol-2016 was close to minor axis and also close to area. Pakistan-2013 was close to area whereas C-273, Barsat, Galaxy-2013, Punjab-96 and Markz-2019 was not close to height, Feret and perimeter but cultivars are showing importance toward these variables. Pirsabak-2004 was close solidity, Ahsan-2016 was close to circularity, Bathor and Maxipak was close to solidity whereas T9 and Wafaq-2000 was away from solidity but still these cultivars having solidity variables. The Figure 5 (b) shows that under well-watered conditions (WW), 41.8 percent of the total variations were accounted for, followed by 25.5 percent of the total variation. The direction of variables was on three sides except one side. Some cultivars was closed to a variables and some are away from variables. For example, Inqilab-91 was close to round variable, NARC-2009 was close to solidity, NARC-2011 was close to solidity variable, Pothowar-700 was close to circularity variable, Parwaz-94 was close to angle variable, Barani-83 was close to minor variable, Suleman-96 was close to perimeter variable, Bakhar-83 was close to angle variables, AS-2002 was close to solidity variable, GA-2002 was close to solidity variable, Chakwal-50 was close to X, GA-2002 was close to height variable and PARI-73 was close to minor variable. C-271 was away from variables such as circularity. Markz-2019, Auqab-2002 and Gold-2016 are away from variables but still fall in a group of ferret, width and major variables, this shows the importance of variables toward the cultivars. Miraj-2008, MH-97 and Millet-2011 was not close to any of variables. The scree plot of the main components of the analysis in multivariate statistics. The scree plot is used to determine how many factors to keep in the principal components for the principal component analysis. Figures 5(c) and 5(d) depict a scree plot illustrating variation explained by each variable under WL and WW conditions.

### 3.4 Allelic Effect of TaCwi gene on grain morphology

The student t-test was performed at a significance of P < 0.05, as shown, Table 3. The allelic frequency is higher in TaCwi-A1a whereas lower in TaCwi-A1b and are further discussed below. (GL) grain length (GA) grain area (GW) grain weight. Table 3 and Figure 6 shows as; The wheat Cwi-21 and CWI-22 primers were used to identify the allelic states at wheat cell wall invertase (TaCwi-A1a) gene in historical wheat cultivars of Pakistan. The frequency of TaCwi-A1a favoured grain size traits and present in 31 varieties, while TaCwi-A1b was unfavourable for grain size traits and was present in 29 varieties. The allelic effect was significant for grain length, width, area and perimeter, while there was non-significant effect on grain roundness.

### 4. Discussion

This experiment was conducted to optimize the high-throughput phenotyping of historical wheat cultivars from Pakistan. Grain phenotyping based on imaging depend on non-destructive ocular analysis of traits and its important purpose is to distinguish the grain size and shape. It provided exact appearances or pictures of wheat grains. Grain characteristics are important agronomic traits because they have a large impact on market value. Because of small differences in the size of wheat cultivars, a large number of measurements are required to obtain grain size. Manual measurement methods have some drawbacks, such as limited data, low measurement accuracy, and the need to gather a wide variety of shape data. A well-organized, consistent, high-throughput grain phenotyping method was needed to validate genetic analysis (Colombo et al. 2022).

The use of software save time and allows for the easy capture and analysis of images. The analysis of over 64 grains per plant at high-resolution (600dpi, 144 pixels per cm) allowed for the differentiation of kernel lines with very minute difference in shape. Without digital assistance, measuring grain shape attributes become less accurate and time-consuming (Rasheed et al., 2014) determined the precise size and shape of grains. Haghshenas et al. (2022) used image J software to determine the grain size and shape. Seed area, length, width, circularity, solidity are the key components of seed size and shape. We found almost the same results as previous grain phenotyping researchers, but our results are more precise and authentic because in this research project elam visualizer software was used to capture wheat grains images that are more authentic and give more variables related to grain size and shapes. The use of machine vision sensor was a good approach to producing binary images that could be used for image analysis. A total of 64 images were captured and analysed. Wheat grains from each cultivars were placed horizontally with proper distance on LED backlight. Quality digital images need to be taken 60 pixels/mm. all photos are named according to the accession number for the genotype. It took almost 0.01 seconds to 0.04 seconds with an average of 0.03 second to analyse a single image. All images will be cropped using infraview software. This software help to improved contrast and brightness of images and reduce shadowing errors in edge detection. Then all photos are stored in designated folder and files are named as JPEG files. The files open in Image J software. The picture file formatted into 8-bit. The image J modified to colour threshold and prevent from false measurements. The global scale is set at 144 pixels per cm with 1.00 known distance and then click OK. Measurements for each grain comes out. The output result then converted to Microsoft Excel where it is further analysed. 19 different parameters were analysed. Each descriptor analysed at P < 0.001 using SAS version 9.0. Circularity, area, perimeters, round and solidity are the parameters determined by image J. The findings are consistent with those, reported recently by (Zhang et al., 2013 and Rasheed et al., 2014). Similarly, (Brinton et al. 2019; Gegas et al., 2010 and Williams et al., 2013) conducted studies on targeted grain variations affecting grain morphology, and the outcomes are related to our findings.

The statistics of every genotype was considered for analyses. The histogram and biplots were constructed for traits in the Microsoft excel 365. Pearson coefficient of correlation was computed using R version 4.0.3. The principal component analysis was carried out that is based on the multivariate phenotypic data. Multivariate analysis was conducted to assess 64 historical wheat cultivars based on well-watered and water limited crops using principal component analysis. Since there was a positive correlation between all variables, the direction of all variables on each side of the plot. Some cultivars have been closed to variables that show importance for those previous traits. The results are very similar to those of previous studies. The phenotypic data for grain morphology, descriptors were averaged from two traits grown under WW and WL conditions. The statistics for measured grain size and shape traits detected in cultivars, as well as the frequency distribution for these traits are shown in Table 4.1 and 4.2. The average grain area under water limited conditions was 0.20 with a range of 0.15 and 0.24cm² minimum grain area was shown by T9 and maximum grain area was shown by Pirsabak-2013. The average grain perimeter under well-watered conditions was 2.07 with a range of 1.71 and 2.35cm². Maximum grain perimeter was recorded for PARI-73 and minimum grain perimeter was shown
by C271 respectively. The average grain circularity under water limited conditions was 0.65 with a range of 0.59 and 0.74cm². Maximum grain circularity was showed by Wafaq-2004 and minimum grain circularity was showed by Barsat. The average grain angle under well-watered conditions was 91.9 with a range of 81.5 and 102.9cm². Minimum grain angle was shown by NARC-2009 and maximum was showed by Punjab-2011 respectively.

For size and shapes traits, a great number of significant correlations have been observed. We've only deliberated important relationships that provide novel vision into the complicated comparison of grain size and shape components for accuracy. There was a significant higher correlation between grain area and grain width (r = 0.86), as well as a strong correlation among grain height and area (r = 0.85). While grain area and circularity had a negative correlation (r = -0.86). Grain length or height and width had a significant positive correlation, indicating that some cultivars may have wider and longer grains at the same time. Grain width, rather than grain length, had a greater positive impact on wheat cultivars. Despite the fact that previous studies found a moderate correlation between grain width and grain height at P = 0.01 (Yu et al 2021 and Dhokia et al., 2003), our findings agree with (Luo et al., 2019) who found a solid correlation among grain weight and size. All previous studies, found grain size and shape to be independent traits in primitive and improved wheat germplasm (Gegas et al., 2010), which is consistent with the findings of this study. TaCwi-A1, have been linked to grain size and weight (Nadolska Orczyk et al., 2017) using comparative genomics, TaGw2 (Su et al., 2011 and Sehgal et al., 2019) was cloned in wheat. Allelic variation of efficient gene in wheat cultivars has been reported in several recent studies (Rasheed et al., 2019). Wheat grain size and weight genes, as well as grain weight GW2 and cell wall invertase CW1, have been developed and validated for wheat cultivars to assess their allelic effects (Su et al., 2011). Student t-test was used to perform association analysis of the significance of P < 0.05. The allelic effect of |TaCwi-A1 gene on grain morphology parameters is frequency, area, grain length, grain width and roundness. The frequency of TaCwi-A1a was present in 31 varieties, of preferred grain traits while TaCwi-A1b was present in 29 varieties unfavourable for grain size traits. The allelic effect was significant for grain length, width, area and perimeter, while there was non-significant effect on grain roundness. The results confirmed that, TaCwi-A1 is an important gene related to grain weight.

Conclusion

It is concluded from the present investigation that the reliability and consistency of digital imaging method is much higher than other methods. The cultivars used in this study showed greater variations in size and shape of important grains under water limited and well-watered conditions. Area, Perimeter, Width, Height, Major, Minor, Angle, Circularity, Ferret, Round and Solidity were recorded. A positive correlation has been found between them. The whole result shows higher accuracy. Using digital imaging many aspects of grain size, shape and weight can be measured and traced in easiest way and in adequate time. So digital imaging technique is highly recommended for reliable and consistent grain phenotyping.

Declarations

**Acknowledgement:** We the authors are grateful for the financial and moral support of Quaid-i-azam University Islamabad, Mohi-ud-Din Islamic University Nerian Sharif and College of Geography and Environment, Henan University Kaifeng for the Research work

**Declaration of competing interest**

The authors of this article declare that they have no conflict of interest.

**Author’s Contribution**

Tehreem Tahir, Asim Shahzad, has equal contribution in designing research, performed the experiments investigation and writing manuscript, The author Awais Rasheed helped in designing research, experiments, and writing the manuscript, The author Sadaf Kayani helped in in revision and editing of the manuscript, all authors have read and approved the final version of the manuscript.

**Data availability statements**

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

**References**


**Tables**

Table 1  64 Historical wheat cultivars of Pakistan
Table 2 (a) Summary statistics of grains size and shape phenotype in Pakistani historical wheat cultivars under water limited conditions using digital imaging technique.
### Table 2 (b) Summary statistics of grains size and shape phenotypes in Pakistani historical wheat cultivars under well-watered conditions using digital imaging technique.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>XM</th>
<th>Y.M.</th>
<th>Perim.</th>
<th>B.X.</th>
<th>B.Y.</th>
<th>Width</th>
<th>Height</th>
<th>Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.15</td>
<td>5.95</td>
<td>5.20</td>
<td>5.96</td>
<td>5.20</td>
<td>1.71</td>
<td>5.67</td>
<td>4.90</td>
<td>0.47</td>
<td>0.491</td>
</tr>
<tr>
<td>Max</td>
<td>0.24</td>
<td>15.3</td>
<td>11.1</td>
<td>15.3</td>
<td>11.1</td>
<td>2.35</td>
<td>15.0</td>
<td>10.8</td>
<td>0.63</td>
<td>0.634</td>
</tr>
<tr>
<td>Range</td>
<td>0.08</td>
<td>9.33</td>
<td>5.95</td>
<td>9.33</td>
<td>5.95</td>
<td>0.64</td>
<td>9.33</td>
<td>5.98</td>
<td>0.15</td>
<td>0.142</td>
</tr>
<tr>
<td>Med.</td>
<td>0.20</td>
<td>11.3</td>
<td>9.20</td>
<td>11.3</td>
<td>9.20</td>
<td>2.06</td>
<td>11.1</td>
<td>8.92</td>
<td>0.56</td>
<td>0.577</td>
</tr>
<tr>
<td>Mean</td>
<td>0.20</td>
<td>11.4</td>
<td>9.073</td>
<td>11.4</td>
<td>9.07</td>
<td>2.07</td>
<td>11.1</td>
<td>8.78</td>
<td>0.56</td>
<td>0.570</td>
</tr>
<tr>
<td>SE.mean</td>
<td>0.002</td>
<td>0.29</td>
<td>0.21</td>
<td>0.21</td>
<td>0.02</td>
<td>0.29</td>
<td>0.21</td>
<td>0.04</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>Var.</td>
<td>0.003</td>
<td>4.26</td>
<td>2.26</td>
<td>4.26</td>
<td>2.12</td>
<td>0.021</td>
<td>4.25</td>
<td>2.14</td>
<td>0.009</td>
<td>0.001</td>
</tr>
<tr>
<td>Std.dev</td>
<td>0.019</td>
<td>2.06</td>
<td>1.45</td>
<td>2.06</td>
<td>1.45</td>
<td>0.147</td>
<td>2.06</td>
<td>1.46</td>
<td>0.307</td>
<td>0.03</td>
</tr>
<tr>
<td>Coef.var</td>
<td>9.3%</td>
<td>18%</td>
<td>16%</td>
<td>18%</td>
<td>16%</td>
<td>7%</td>
<td>18%</td>
<td>16%</td>
<td>5.4%</td>
<td>6%</td>
</tr>
</tbody>
</table>

### Table 3 Allelic effect of the TaCwi-A1 gene on various grain morphology parameters in Pakistani historical wheat cultivars.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
<th>Area (cm)</th>
<th>Grain Length (cm)</th>
<th>Grain Width (cm)</th>
<th>Roundness</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaCwi-A1a</td>
<td>31</td>
<td>0.22^a</td>
<td>0.73^a</td>
<td>0.36^a</td>
<td>0.502ns</td>
</tr>
<tr>
<td>TaCwi-A1b</td>
<td>29</td>
<td>0.19^b</td>
<td>0.68^b</td>
<td>0.34^b</td>
<td>0.503ns</td>
</tr>
</tbody>
</table>

Students t-test Significance at P < 0.05

### Figures
Figure 1

Steps involved in digital image analysis of historical wheat cultivars

(A)  

(B)
Figure 2

Historical Wheat cultivars of Pakistan. (A) right and (B) left shows the image of grains on scanner.

Figure 3

(a) Box plots showing variation in important grains size and shape characteristics in 64 wheat cultivars under (WW) conditions and (WL) conditions using digital imaging technique.

(b) Stack Histogram showing variation in for important grains characteristics in 64 wheat cultivars under (WW) and (WL) conditions using digital imaging technique.

Figure 4

Coefficient of correlation between grain size and shapes traits of 64 historical wheat cultivars under (WW) and (WL) conditions using imaging analysis.
Figure 5

(a) PCA biplot under water limited (WL) condition; showing grain variable as a vectors and wheat cultivars as points in the PCA biplot of 64 historical wheat cultivars of Pakistan.

(b) PCA biplot under well-watered (WW) conditions showing grain variable as a vectors and wheat cultivars as a point in the PCA biplot of 64 historical wheat cultivars

(c) Scree plot showing the variation explained by each variable under water limited (WL) conditions of 64 historical wheat cultivars of Pakistan.

(d) Scree plot showing the variation explained by each variable under well-watered (WW) conditions of 64 historical wheat cultivars of Pakistan.
Figure 6

Box plots showing Allelic effects of TaCwi-A1 gene on (a) GL (b) GA (c) Roundness (d) GW