

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Theses, Dissertations, and Student Research in  
Agronomy and Horticulture

Agronomy and Horticulture Department

---


11-2018

# Identification of Genes/Genomic Regions Controlling Resistance to Biotic and Abiotic Stresses in Synthetic Hexaploid Wheat

Madhav Bhatta

*University of Nebraska-Lincoln*

Follow this and additional works at: <http://digitalcommons.unl.edu/agronhortdiss>

 Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

---

Bhatta, Madhav, "Identification of Genes/Genomic Regions Controlling Resistance to Biotic and Abiotic Stresses in Synthetic Hexaploid Wheat" (2018). *Theses, Dissertations, and Student Research in Agronomy and Horticulture*. 156.  
<http://digitalcommons.unl.edu/agronhortdiss/156>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Theses, Dissertations, and Student Research in Agronomy and Horticulture by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

IDENTIFICATION OF GENES/GENOMIC REGIONS CONTROLLING RESISTANCE TO BIOTIC  
AND ABIOTIC STRESSES IN SYNTHETIC HEXAPLOID WHEAT

by

Madhav Bhatta

A DISSERTATION

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment of Requirements  
For the Degree of Doctor of Philosophy

Major: Agronomy and Horticulture  
(Plant Breeding and Genetics)

Under the Supervision of Professor P. Stephen Baenziger

Lincoln, Nebraska

November, 2018

IDENTIFICATION OF GENES/GENOMIC REGIONS CONTROLLING RESISTANCE TO  
BIOTIC AND ABIOTIC STRESSES IN SYNTHETIC HEXAPLOID WHEAT

Madhav Bhatta, Ph.D.

University of Nebraska, 2018

Advisor: P. Stephen Baenziger

Synthetic hexaploid wheat (SHW;  $2n=6x=42$ , AABBDD, *Triticum aestivum* L.) is produced from an interspecific cross between durum wheat ( $2n=4x=28$ , AABB, *T. turgidum* L.) and goat grass ( $2n=2x=14$ , DD, *Aegilops tauschii* Coss.). It is reported to have a considerable amount of genetic diversity and is a potential source of novel alleles controlling abiotic and biotic stresses resistance and improving wheat quality. Therefore, the first study was to understand the genetic diversity and population structure of SHWs and compare the genetic diversity of SHWs with elite bread wheat (BW) cultivars. The result of this study identified a wide range of genetic diversity within the SHWs. The genetic diversity of the ABD and D-genome of SHWs were 50% and 88.2%, respectively, higher than that found on the respective genome in a sample of elite BW cultivars. The second study was to identify novel genomic regions and underlying genes associated with grain yield and yield-related traits under two drought-stressed environments. This study identified 90 novel genomic regions and haplotype blocks associated with improving grain yield and yield-related traits with phenotypic variance explained of up to 32.3%. The third study was to identify common bunt resistance genotypes, genomic regions and underlying genes conferring resistance to common bunt. This study identified 29 resistant SHWs and 15 genomic regions (five were novel)

conferring resistance to common bunt. The fourth study to explore the genetic variation of 10-grain minerals (Ca, Cd, Co, Cu, Fe, Li, Mg, Mn, Ni, and Zn) and grain protein concentration (GPC); identify marker-trait associations and candidate genes associated with grain minerals using a genome-wide association study (GWAS). A wide range of genetic variation identified within SHWs for GPC and grain mineral concentrations. A GWAS identified 92 genomic regions (60 were novel and 40 were within genes) associated with increasing beneficial grain mineral concentration and decreasing concentration of toxic compound such as Cd. The results from this research will be valuable for broadening the genetic base of wheat and could assist in further understanding of the genetic architecture of traits under biotic and abiotic stresses.

## Acknowledgments

I would like to express my deep appreciation to Dr. P. Stephen Baenziger for being my major advisor and Dr. Alexey Morgounov for being my collaborating-advisor. This work would not have been accomplished without their guidance, suggestions, and encouragements throughout the journey.

My sincere gratitude to my statistics minor advisor, Dr. Kent M. Eskridge, for motivating and guiding me throughout the journey and assisting in solving the problems related to statistics. My gratitude is also extended to Dr. Teshome Regassa who motivated in every step of this journey and served on the supervisory committee. I am also thankful to Dr. Stephen Wegulo who guided and supported me in the wheat fungal diseases related experiments and served on the supervisory committee. All the supervisory committee member's support, comments, and suggestions during the research and dissertation writing were also invaluable to accomplish this task.

I would also like to acknowledge the Monsanto Beachell-Borlaug International Scholarship Program, University of Nebraska-Lincoln, Lincoln, USA, and International Maize and Wheat Improvement Center (CIMMYT) at Turkey for their support throughout the project; to the IWWIP for providing seeds of synthetic hexaploid wheats; to the CIMMYT-Turkey staffs especially Adem Urglu and Ibrahim Ozturk; to Dr. Brian Waters for providing support in Mineral digestion; to Dr. Javier Seravali at the University of Nebraska Redox Biology Center Proteomics and Metabolomics Core for conducting the mineral analysis (ICP-MS); to the Bahri Dagdas International Agricultural Research Institute staffs including Emel Ozer, Fatih Ozdemir, and Enes Yakisir; to Seregey Schepelew from Omsk, Russia and Akmal Meyliev from

Uzbekistan for assisting in field data collections in 2017; to Dr. Jesse Poland for providing the genotyping-by-sequencing (GBS) data; to Shuangye Wu for technical support for GBS data and Sarah Blecha for assisting in DNA extraction and sending samples for genotyping; to the Holland Computing Center at the University of Nebraska, Lincoln, for providing the high performance computing resources; to Dr. Vikas Belamkar for his valuable suggestions throughout the project; to Mitch Montgomery and Greg Dorn for assisting in the greenhouse and field experiments; to all technical staffs, graduate, and undergraduate students in the Dr. Baenziger's laboratory for their valuable suggestions; and to professors, staffs, and friends at the Department of Agronomy and Horticulture for their help during the entire course of study.

Last, but not the least, a special acknowledgment goes to my wife (Rachana Poudel) for giving me strength, patience, support, and proofreading this dissertation. I would also add special thanks to my dad (Krishna Prasad), mom (Dharma Devi), sister (Mina), brothers (Tej and Naresh), and relatives for their continual support and motivations throughout the life.

## PREFACE

Synthetic hexaploid wheat is made by crossing modern durum wheat and wild goat grass. It has been reported to be an efficient and beneficial resource for broadening the genetic base of bread wheat. This dissertation was focused on understanding the genetic diversity of unique sets of winter synthetic wheat germplasm and unlocking their genomic regions for controlling the biotic and abiotic stress resistance and for improving wheat nutritional quality.

### Organization of the dissertation

The Chapters 2, 3, 4, and 5 presented in the dissertation have been published in the *BMC genomics* (Bhatta M., Morgounov A., Belamkar V., Poland J., Baenziger P.S., *BMC Genomics* 19:591 (2018). <https://doi.org/10.1186/s12864-018-4969-2>), *International Journal of Molecular Sciences* (Bhatta M., Morgounov A., Belamkar V., Baenziger P.S. *International Journal of Molecular Sciences* 19:3011 (2018). <https://doi.org/10.3390/ijms19103011>), *Euphytica* (Bhatta M., Morgounov A., Belamkar V., Yorgancilar A., Baenziger P.S. *Euphytica* 214:200 (2018). <https://doi.org/10.1007/s10681-018-2282-4>), and *International Journal of Molecular Sciences* (Bhatta M., Baenziger P.S., Waters B.M., Poudel R., Belamkar V., Poland J., Morgounov A., *International Journal of Molecular Sciences* 19(10):3237 (2018). <https://doi.org/10.3390/ijms19103237>), respectively. Therefore, these chapters have been formatted using authors guidelines from the respective Journals.

## CHAPTER 1 Table of Contents

LISTS OF MULTIMEDIA OBJECTS .....	vii
CHAPTER 1. LITERATURE REVIEW .....	1
Introduction .....	1
Evolution of common bread wheat .....	1
Production of synthetic hexaploid wheat .....	2
Synthetic hexaploid wheat for improving the genetic diversity of bread wheat .....	2
Synthetic hexaploid wheat as a potential source of biotic and abiotic stresses resistance .....	4
Synthetic hexaploid wheat as a potential source for improving grain protein and mineral concentrations .....	6
Goals and objectives .....	6
References .....	9
FIGURES .....	15
CHAPTER 2. UNLOCKING THE NOVEL GENETIC DIVERSITY AND POPULATION STRUCTURE OF SYNTHETIC HEXAPLOID WHEAT .....	17
ABSTRACT .....	17
INTRODUCTION .....	18
METHODS .....	20
Plant Material .....	20
Genotyping and SNP Discovery .....	21
Genetic Diversity and Population Structure Analysis .....	21
RESULTS .....	22
Population Structure .....	23
Genetic Diversity between the Two Synthetic Hexaploid Wheat Groups .....	25
Genetic Diversity of Synthetic Hexaploid Wheat and Bread Wheat Cultivars .....	26
DISCUSSION .....	27
Population Structure .....	27
Genetic Diversity .....	29
CONCLUSIONS .....	30
REFERENCES .....	32
TABLES .....	36
FIGURES .....	41



CHAPTER 3. GENOME-WIDE ASSOCIATION STUDY REVEALS NOVEL GENOMIC REGIONS FOR GRAIN YIELD AND YIELD-RELATED TRAITS IN DROUGHT-STRESSED SYNTHETIC HEXAPLOID WHEAT .....	45
ABSTRACT .....	45
Introduction .....	46
Materials and Methods .....	48
Site Description .....	48
Plant Materials and Experimental Design.....	48
Trait Measurements .....	49
Phenotypic Data Analysis.....	49
Genotyping and SNP Discovery .....	52
Population Structure and Genome-Wide Association Study Analysis .....	53
Haplotype Block Analysis .....	54
Putative Candidate Gene Analysis.....	54
Results and Discussion .....	55
Weather Conditions .....	55
Phenotypic Variation for Yield and Yield-Related Traits.....	55
Principal Component Analysis and Phenotypic Correlation.....	56
Population Structure and Genome-Wide Association Study .....	57
Potential Candidate Gene Annotations Affecting Yield and Yield-Related Traits Under Drought Stress.....	69
Conclusions .....	69
References .....	72
TABLES.....	87
FIGURES .....	91
CHAPTER 4. GENOME-WIDE ASSOCIATION STUDY REVEALS FAVORABLE ALLELES ASSOCIATED WITH COMMON BUNT RESISTANCE IN SYNTHETIC HEXAPLOID WHEAT .....	95
Abstract.....	95
Introduction .....	97
Materials and Methods .....	99
Site description and plant materials .....	99
Disease evaluation and experimental design .....	99
Phenotypic data analysis.....	100
Genotyping and SNP Discovery .....	101

Population structure, genome-wide association analysis, and candidate gene analysis .....	101
Results and Discussion .....	102
Phenotypic distribution of common bunt incidence in the SHW population .....	102
Genome-wide association study .....	103
Potential candidate genes underlying marker-trait associations and their gene annotations .....	104
Favorable alleles associated with common bunt resistance .....	106
Conclusions .....	106
References .....	108
TABLES .....	114
Figures .....	115
CHAPTER 5. A GENOME-WIDE ASSOCIATION STUDY REVEALS NOVEL GENOMIC REGIONS ASSOCIATED WITH 10 GRAIN MINERALS IN SYNTHETIC HEXAPLOID WHEAT .....	120
Introduction .....	121
Materials and methods .....	123
Plant materials and experimental design.....	123
Grain yield, thousand kernel weight, grain protein concentration, and grain mineral analysis.....	123
Phenotypic data analysis.....	124
Genotyping and SNP discovery .....	125
Population structure and genome-wide association analysis .....	126
Results and Discussion .....	126
Phenotypic variation for grain protein concentration and grain minerals .....	126
Principal component analysis and phenotypic correlation .....	128
Selection of top-ranking genotypes .....	130
Population structure and genome-wide association study.....	131
<i>Cadmium</i> .....	132
Relationship between grain mineral concentrations and number of favorable alleles .....	135
Multi-trait and stable marker-trait associations .....	136
Gene underlying marker-trait associations .....	137
Conclusions .....	138
References .....	141
TABLES .....	151
FIGURES .....	154
APPENDIX .....	157

**APPENDIX VIII.** Number of synthetic hexaploid wheat germplasm having either favorable or unfavorable alleles associated with common bunt resistance. .... 178

## LISTS OF MULTIMEDIA OBJECTS

### CHAPTER 1. LITERATURE REVIEW

Figure 1. Bread wheat evolution

Figure 2. Production of synthetic hexaploid wheat

### CHAPTER 2. UNLOCKING THE NOVEL GENETIC DIVERSITY AND POPULATION STRUCTURE OF SYNTHETIC HEXAPLOID WHEAT

Table 1. Distribution of SNP markers and genetic diversity summary statistics of 101 synthetic hexaploid wheats including observed nucleotide diversity ( $\pi$  bp<sup>-1</sup>), expected nucleotide diversity ( $\theta$  bp<sup>-1</sup>), Tajima's D, effective number of alleles (Eff-num), observed heterozygosity ( $H_o$ ), and gene diversity ( $H_s$ ).

Table 2. Population genetic diversity summary statistics of two subgroups in 101 synthetic hexaploid wheats (SHWs), Durum wheat, and *Aegilops tauschii* (*Aegilops*) obtained from GenoDive including effective number of alleles (Eff-num), observed heterozygosity ( $H_o$ ), and gene diversity ( $H_s$ ), and Mean  $F_{ST}$  obtained from STRUCTURE.

Table 3. A standardized measure of population differentiation ( $F'_{ST}$ ), Jost's D as an index of population differentiation, and Nei's D as the standard genetic distance in two subgroups in 101 synthetic hexaploid wheats (SHWs) was computed in GenoDive.

Table 4. Analysis of molecular variance (AMOVA) within and between the two subgroups of 101 synthetic hexaploid wheats identified by the Bayesian clustering.

Table 5. Population genetic diversity summary statistics of 101 synthetic hexaploid wheats (SHWs) and 12 bread wheat cultivars obtained from GenoDive.

Figure 1. Distribution of 35,939 single nucleotide polymorphisms (SNPs) across 21 chromosomes and unanchored scaffolds in 101 synthetic hexaploid wheats.

Figure 2. Population structure of the 101 synthetic hexaploid wheat germplasm. A: Line graph of delta K over K from 1 to 10, and the highest peak was observed at Delta K=2, suggesting the synthetic hexaploid wheat (SHW) germplasm has two subgroups. B: The two subgroups identified from the STRUCTURE and grouped based on the geographical location of the durum parents and growth habit of the crop. C: Cluster analysis (neighbor joining) and D: Principal coordinate analysis (PCoA). Color reflects grouping derived from STRUCTURE.

Figure 3. Population structure of the durum parents used in the production of synthetic hexaploid wheat (SHW) germplasm. A: Line graph of delta K over K from 1 to 10, and the highest peak was observed at Delta K=2, suggesting the durum wheat used in this study has two subgroups. B: The two subgroups were identified from the STRUCTURE and grouped based on the

geographical location of the durum parents and growth habit of the crop. C: Cluster analysis (neighbor joining) and D: Principal coordinate analysis (PCoA). Color reflects grouping derived from STRUCTURE.

Figure 4. Population structure of the *Aegilops* parents used in the production of synthetic hexaploid wheat. A: Line graph of delta K over K from 1 to 10, and the highest peak was observed at Delta K=2, suggesting the *Aegilops* used in this study has two subgroups. B: The two subgroups were identified from the STRUCTURE and grouped based on the type of *Aegilops* parents used. C: Cluster analysis (neighbor joining) and D: Principal coordinate analysis (PCoA). Color reflects grouping derived from STRUCTURE.

### CHAPTER 3. GENOME-WIDE ASSOCIATION STUDY REVEALS NOVEL GENOMIC REGIONS FOR GRAIN YIELD AND YIELD-RELATED TRAITS IN DROUGHT-STRESSED SYNTHETIC HEXAPLOID WHEAT

Table 1. Mean monthly temperatures and total monthly rainfalls in two growing seasons (2016 and 2017) and 25-year average data in Konya, Turkey.

Table 2. Phenotypic variation for grain yield and yield-related traits with least squares means, range, coefficient of variation (CV), and broad sense heritability ( $H^2$ ) of 123 drought-stressed synthetic hexaploid wheat grown in two seasons (2016 and 2017) in Konya, Turkey.

Table 3. List of significant markers associated with grain yield and yield-related traits, favorable alleles (underlined), SNP effects, and drought-related putative genes from genome-wide association study of 123 drought stressed synthetic hexaploid wheat grown in 2016 in Konya, Turkey.

Table 4. List of significant markers associated with grain yield and yield-related traits, favorable alleles (underlined), SNP effects, and drought-related putative genes obtained from genome-wide association study of 123 drought stressed synthetic hexaploid wheat grown in 2017 in Konya, Turkey.

Figure 1. Principal component bi-plot analysis of 123 drought-stressed synthetic hexaploid wheat grown in two seasons (A: 2016 and B: 2017) in Konya, Turkey. AWLN, awn length; BMWT, biomass weight; FLA, flag leaf area; FLLN, flag leaf length; FLW, flag leaf width; GVWT, grain volume weight; GY, grain yield; HI, harvest index; RTLN, root length; STMDIA, stem diameter; and TKW, thousand kernel weight.

Figure 2. Significant marker trait associations identified on each chromosome for grain yield and yield-related traits obtained from genome-wide association study of 123 synthetic hexaploid wheat grown in 2016 and 2017 in Konya, Turkey.

Figure 3. Linkage disequilibrium (LD) values ( $R^2$ ) and haplotype blocks with significant marker-trait associations (MTAs;  $\geq 2$ ) observed (A) on chromosome 7A for grain yield (B) on chromosome 3A for grain yield (C) on chromosome 3A for biomass weight (D) on chromosome 3B for stem diameter (E) on chromosome 1A for flag leaf area (F) on chromosome 6B for flag leaf area (G) on chromosome 7D for flag leaf area (H) on chromosome 6D for root length and

phenotypic variance explained (PVE) by each haplotype block. Dark red color represents the strong LD whereas light red color represents the weak LD between pairs of MTAs.

Figure 4. Potential candidate gene functions harboring SNPs affecting yield and yield-related traits under drought stress. The count of marker-trait associations (for either single or multiple traits) located within genes that have the same gene annotation is shown. AWLN, awn length; BMWT, biomass weight; FLA, flag leaf area; FLLN, flag leaf length; FLW, flag leaf width; GVWT, grain volume weight; GY, grain yield; HI, harvest index; RTLN, root length; STMDIA, stem diameter; and TKW, thousand kernel weight.

#### CHAPTER 4. GENOME-WIDE ASSOCIATION STUDY REVEALS FAVORABLE ALLELES ASSOCIATED WITH COMMON BUNT RESISTANCE IN SYNTHETIC HEXAPLOID WHEAT

Table 1. List of significant markers associated with common bunt resistance, favorable allele (underlined), SNP effects, and functional annotation of genes containing or flanking significant SNPs from the genome-wide association study of 125 synthetic hexaploid wheats grown in 2016 and 2017 in Eskisehir, Turkey.

Fig. 1 Frequency distribution of common bunt infected spikes (%) obtained from best linear unbiased predictors combined over two years (2016 and 2017) from 125 synthetic hexaploid wheats.

Fig. 2 Physical distribution of 35,798 genotyping-by-sequencing derived SNPs within 1-Mb window size on 21 chromosomes of 125 synthetic hexaploid wheats.

Fig. 3 Manhattan (main panel) and quantile-quantile (Q-Q; top right) plots showing genome-wide association results for common bunt resistance in 125 synthetic hexaploid wheat lines. The Manhattan plot shows the association  $-\log_{10}(p)$  for each genome-wide SNPs (35,798) on y-axis by chromosomal position on x-axis. The green dots in the Manhattan plot shows the significant marker-trait associations above the threshold line (red) with  $P = 8.8E-05$  [ $-\log_{10}P=4.05$ ]. The Q-Q plot shows the deviation of association test statistics (blue dots) from the distribution expected under the null hypothesis (red line).

Fig. 4 Distribution of favorable and unfavorable alleles in 125 synthetic hexaploid wheat lines and comparison of allelic for the SNP markers associated with common bunt incidence (%) to determine the significant differences between the mean values of two alleles. The cross symbol and dividing line inside the box plots are mean and median values of common bunt incidence, respectively. \*, \*\*\* Indicate significance at the 0.05 and 0.001 probability levels, respectively.

Fig. 5 Scatter plots and regression analysis of common bunt incidence (%) and the number of (a) favorable alleles or (b) unfavorable alleles in 125 synthetic hexaploid wheat lines.

## CHAPTER 5. A GENOME-WIDE ASSOCIATION STUDY REVEALS NOVEL GENOMIC REGIONS ASSOCIATED WITH 10-GRAIN MINERALS IN SYNTHETIC HEXAPLOID WHEAT

Table 1. Analysis of variance and phenotype variation for 10 grain minerals, grain protein content and grain yield with minimum (min), maximum (max), fold change (max/min), mean, and broad sense heritability ( $H^2$ ) values of 123 synthetic hexaploid wheats grown in 2016 and 2017 growing seasons in Konya, Turkey.

Table 2. Pearson's correlation coefficients of 10 grain minerals, grain protein content (GPC), grain yield (GY), and GY controlling for thousand kernel weight ( $GY_{pTKW}$ ) in 123 synthetic hexaploid wheat grown in 2016 (upper triangle) and 2017 (lower triangle) growing seasons in Konya, Turkey.

Table 3. Potential candidate genes containing/flanking marker-trait associations for improving grain minerals in SHWs.

Figure 1. Factor analysis using principal component method based on correlation matrix on grain yield, grain protein concentration, and 10 grain mineral concentrations in 123 synthetic hexaploid wheat lines grown in 2016 (A) and 2017 (B) in Konya, Turkey.

Figure 2. Significant marker-trait associations identified on each chromosome for 10 grain minerals from a genome-wide association study using 35,648 SNPs in 123 synthetic hexaploid wheat grown in 2016 and 2017 in Konya, Turkey.

Figure 3. Regression analysis between number of favorable alleles and best linear unbiased predictor values of grain minerals concentrations obtained from two years (2016 and 2017) experiments conducted in Konya, Turkey. Favorable alleles are defined as the alleles that increases the grain concentration of beneficial minerals such as calcium (A), cobalt (C), copper (D), Iron (E), Lithium (F), Magnesium (G), Manganese (H), Nickel (I), and Zinc (J), whereas decreases grain cadmium (B) concentration.

## CHAPTER 1. LITERATURE REVIEW

### Introduction

Wheat is one of the most widely grown cereal crops with the production of more than 756 million tons in 2017-18 [1] and it feeds more than one-third of the world's population [2]. Wheat grain is a good source of carbohydrate, protein, antioxidants, fiber, and minerals and it supplies nearly 20% of the food calories in our diet [2]. Wheat per capita consumption was estimated at 60.4 and 94.9 kg in 2012-2014 in developing and developed countries, respectively, with an average of 67.2 Kg in the world [3]. The world consumption rate of wheat will increase with an increase in the world population and prosperity. It is estimated that wheat production should be increased by 1.34 and 0.65% in developing and developed countries, respectively, with an average increase by 1.09% to feed the world by 2024 [3]. However, wheat yields have plateaued over the last 15 years [4] and the wheat productivity is projected to decline up to 8% (up to 25% in some tropical regions) due to climate change [5]. To meet the global food demand of 9.6 billion (estimated) population by 2050 [6], wheat breeders have tremendous challenges to increase grain yield and yield stability, improve end-use and nutritional quality characteristics, increase resistance to multiple biotic and abiotic stresses, and ultimately increase the rate of genetic gain under rapidly changing climate [7].

### Evolution of common bread wheat

Modern bread wheat (*Triticum aestivum* L.) evolved from a natural hybridization between the tetraploid cultivated emmer wheat [*T. turgidum* L. spp. *diccicum* (Schrank) Thell.; a

progenitor of durum wheat] and the wild diploid goat grass (*Ae. tauschii* Coss.) about 8,000 years ago in the Fertile Crescent [8-10] (Figure 1). The cultivated emmer wheat was produced during the process of domestication and selection of wild emmer wheat (*T. dicoccoides*) [11]. The wild emmer wheat was evolved from a natural cross between an unknown diploid grass species with the closest relative being *Ae. speltooides* (BB;  $2n=2x=14$ ) and a diploid wild grass species, *T. urartu* (AA;  $2n=2x=14$ ) [8,11] (Figure 1).

### **Production of synthetic hexaploid wheat**

Synthetic hexaploid wheat is made by crossing modern durum wheat (*T. turgidum* L.) and wild goat grass (*Ae. tauschii* Coss.) (Figure 2) [12]. Embryo rescue is performed following crossing to save the embryos which have the genomic constitution of ABD genomes [13]. As this form is amphiploid, sterile, and unstable, the chromosomes are doubled using colchicine to form a stable hexaploid wheat, commonly referred as synthetic hexaploid wheat (Figure 2) [14]. Synthetic hexaploid wheat is easily crossable with elite bread wheat cultivars because they have similar floral attributes and the same genomic constitution [14].

### **Synthetic hexaploid wheat for improving the genetic diversity of bread wheat**

Genetic diversity is the foundation of any crop improvements. Several studies have found the genetic diversity of wheat is lower compared to their progenitors [15-17]. The bottleneck for wheat genetic diversity relates to the recent origin of bread wheat, around 8,000 years ago [18], where few crosses between tetraploid and diploid progenitors were assumed to be involved in the production of bread wheat. Also, most breeding programs rely on a small number of parents in



developing germplasm pools which ultimately results in even narrower genetic diversity [19]. Such narrow genetic diversity of elite wheat germplasm is a challenge for sustainable wheat production with the changing climate and rapidly growing world population.

However, and while unintended, recurrent selection in wheat improvement also resulted in the loss of potentially valuable genetic diversity in creating elite cultivars. A large amount of genetic diversity or favorable alleles for biotic and abiotic stresses resistance, yield and yield-related traits, and quality traits have been preserved in crop wild relatives [20]. Many plant breeders have recovered or improved genetic diversity by crossing their elite lines with landraces or wild relatives or through the production of synthetic hexaploid wheat [12,20,21].

Over the past few decades, synthetic hexaploid wheat has given priority in many breeding programs especially in the International Maize and Wheat Improvement Center (CIMMYT) for broadening the genetic base of bread wheat [12,15,18,21]. The CIMMYT has developed more than 1000 SHWs from crosses of more than 600 *Ae. tauschii* accessions [12]. The SHWs were found to have a novel source of genetic diversity especially in the D-genome [16,17,21]. Several studies have identified the higher genetic diversity of SHWs compared to bread wheat cultivars. For example, a study on 101 SHWs using ~36,000 genotyping-by sequencing (GBS)-derived SNP markers had identified that the genetic diversity of SHW was twice the genetic diversity of elite bread wheat cultivars [21]. Similarly, higher genetic diversity in SHWs have been reported in past using different markers systems such as amplified fragment length polymorphism (AFLP) [16] and short sequence repeat (SSR) markers [17]. Hence, the wide range of genetic diversity exists in SHWs that could potentially be utilized in a wheat breeding program for bringing both novel alleles that have never been exploited, as well as alleles that have been lost during the process of recurrent selection or domestication for mitigating the crop

production challenges such as biotic and abiotic stresses under rapidly changing climatic conditions.

### **Synthetic hexaploid wheat as a potential source of biotic and abiotic stresses resistance**

Global climate has been changing rapidly with an estimated global temperature increase up to 6.4 °C [22] and it will impose a wide range of constraints on agricultural production and productivity especially by increasing several biotic and abiotic stresses. Drought is the most important abiotic stress that reduces agricultural production and productivity worldwide [23]. Therefore, breeding for drought tolerance is important for wheat improvement. However, bread wheat has limited genetic and phenotypic diversity available for breeding for drought tolerance [24]. The SHWs are potential sources of new genetic variation for drought tolerance in wheat improvement. Several studies on synthetic derived lines (SDLs) indicated that the SDLs provided up to 45% yield increase compared to their wheat parents under drought stressed conditions [24, 25]. Similarly, SDLs produced up to 30% yield increase compared with parent lines and local checks under rainfed conditions [26]. The synthetic derived cultivar named Chuanmai-42 developed in China was found to have 35% higher grain yield than the commercial check cultivar [27]. Therefore, exploiting SHWs for drought tolerance is needed for the global food security.

Biotic stresses such as diseases and insect-pest infestation are a major constraint to wheat production. Therefore, genetic resistance against biotic stress is a major goal in wheat breeding program. However, modern wheat cultivars have a limited genetic variation for diseases and insect-pest resistance [12] and there is always the possibility of the evolution of new

diseases/insect-pest or races to overcome previously identified resistance genes. Therefore, a wide range of genetic variation is prerequisite for protecting crop productivity and genetic gain. Identification of genetic resistance to multiple diseases and pests is a prerequisite for any breeding programs for the sustainable agricultural productivity and production. Therefore, it is important to study new genetic resources that have the potential to add genetic variation for several biotic and abiotic stresses resistance. This need may be helped by increasing the genetic variation of wheat through the utilization of SHWs [12]. The D-genome from wild goat grass used in the SHWs have shown have many desirable genes for wheat improvement including disease and insect pest resistance [12]. Several studies identified that SHWs are resistance to biotic stresses. For example, SHWs were found to have resistance to leaf rust (incited by *Puccinia triticina*) [28, 29,30], stem rust (incited by *P. graminis*) [29,30], stripe rust (incited by *P. striiformis*) [29,30,31], Fusarium head blight (incited by *Fusarium graminearum*) [28], yellow spot (incited by *Pyrenophora tritici-repentis*) [29,30], Septoria nodorum (incited by *Parastagonospora nodorum*) [29,30], Septoria tritici blotch (incited by *Mycosphaerella graminicola*) [28,29], cereal cyst nematode (incited by *Heterodera avenae*) [29], crown rot (incited by *F. pseudograminearum*) [30], root-lesion nematode (incited by *Pratylenchus thornei* and *P. neglectus*) [29], and Karnal bunt (*Tilletia indica*) [32]. Additionally, SHWs had multiple insect-pest resistance [13,28,30]. Therefore, exploiting genetic variation of SHWs is needed for the genetic improvement of wheat under biotic stress.

## **Synthetic hexaploid wheat as a potential source for improving grain protein and mineral concentrations**

Grain minerals are required in the human diet for their wellbeing and can be supplied through the consumption of a diversified diet. However, people in the developing countries mostly depend on wheat and rice (*Oryza sativa* L.) as a staple crop and suffer from grain mineral deficiencies [33]. It has been estimated that over 30 and 60% of the world's population suffer from Zn and Fe deficiencies, respectively [33]. Other minerals deficiencies such as Ca, Cu, and Mg have been reported in many developed and developing countries [33,34]. Wheat is one of the most consumed cereal crops and the improvement of grain minerals is essential for food and nutritional security. However, the modern wheat cultivars have a low concentration of grain minerals [35] and found to have a narrow genetic variation compared to its wild relatives [36]. The SHW is a potential source of high grain protein [37] and mineral concentrations [38]. However, very limited studies have been conducted to evaluate the genetic variation for grain protein and mineral concentration in SHWs.

### **Goals and objectives**

The importance of biotic and abiotic stress resistance of wheat to ensuring food security in future climate scenarios is not disputed, and the potential of wide-scale use of genetic resources from SHW to accelerate and better focus breeding outcomes is well known. Examples include the presence of SHW in the pedigrees of up to 50% of lines in the CIMMYT international nurseries and the use of physiological trait-based hybridization strategies to incorporate useful genetic variation from genetic resources into elite backgrounds [12,13].

However, the success of SHW utilization in a breeding program could have been much higher if they were guided by the knowledge of genes/genomic regions controlling resistance to biotic and abiotic stresses and increasing grain quality. Therefore, the main goal of this study was to use this rich genetic resource to identify superior primary synthetics possessing resistances to diseases and drought and identify the respective genes/genomic regions that can be used for marker-assisted transfer of the genes into high-yielding modern elite wheat germplasm. Additionally, the second goal was to evaluate the variation within this synthetic wheat germplasm for improved grain quality (especially grain protein concentration) and mineral concentrations and identify the genes/genomic regions contributing to better end-use and nutritional quality. Finally, our goal was to select and utilize the top-ranking genotypes identified as an excellent source for multiple traits of interests into an elite winter wheat breeding program and incorporate novel genomic regions for multiple-traits in a marker-assisted selection method upon validation in an independent population.

Specific objectives of this study were as follows:

1. To investigate genetic diversity in unique sets of diverse SHW accessions using GBS-derived SNPs, decipher the presence of population structure in SHW collection, and compare genetic diversity among SHWs and elite wheat cultivars to determine the prospects of broadening the genetic base of bread wheat using SHW.
2. To identify novel genomic regions associated with grain yield and yield-related traits and identify the underlying genes for the significant genomic regions identified under drought stressed condition and investigate their potential role for drought tolerance using functional annotations.

3. To identify common bunt resistance genotypes and genomic regions associated with common bunt resistance and further investigate the significant SNPs present within genes using the functional annotations of the underlying genes.
4. To explore the genetic variation for grain protein and mineral concentrations, identify marker-trait associations for increasing grain mineral concentrations and investigate potential candidate genes associated with grain mineral concentration using functional annotation.

## References

1. World population prospects 2017.  
<https://esa.un.org/unpd/wpp/Download/Standard/Population/> (accessed Aug 8, 2018).
2. Poudel, R. and Bhatta, M., 2017. Review of nutraceuticals and functional properties of whole wheat. doi: 10.4172/2155-9600.1000571
3. OECD 2015. "Wheat projections: Consumption, food use, per capita", in *OECD-FAO Agricultural Outlook 2015*, OECD Publishing, Paris, [https://doi.org/10.1787/agr\\_outlook-2015-table121-en](https://doi.org/10.1787/agr_outlook-2015-table121-en)
4. Grassini, P., Eskridge, K.M. and Cassman, K.G., 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nature communications*, 4, p.2918.
5. Schleussner, C.F., Lissner, T.K., Fischer, E.M., Wohland, J., Perrette, M., Golly, A., Rogelj, J., Childers, K., Schewe, J., Frieler, K. and Mengel, M., 2016. Differential climate impacts for policy-relevant limits to global warming: the case of 1.5 C and 2 C. *Earth system dynamics*, 7, pp.327-351.
6. World food situation, 2018. <http://www.fao.org/worldfoodsituation/csdb/en/> (accessed Aug 8, 2018).
7. Winfield, M.O., Allen, A.M., Wilkinson, P.A., Burrridge, A.J., Barker, G.L., Coghill, J., Waterfall, C., Wingen, L.U., Griffiths, S. and Edwards, K.J., 2018. High-density genotyping of the AE Watkins Collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat. *Plant biotechnology journal*, 16(1), pp.165-175.

8. Feuillet, C., Langridge, P. and Waugh, R., 2008. Cereal breeding takes a walk on the wild side. *TRENDS in Genetics*, 24(1), pp.24-32.
9. Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R. and Martin, W., 2002. Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetics*, 3(6), p.429.
10. Dubcovsky, J. and Dvorak, J., 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, 316(5833), pp.1862-1866.
11. Feldman, M. 2001 Origin of cultivated wheat, In: Alain P. Bonjean and William J. Angus (Eds.), *The world wheat book: A history of wheat breeding*, Intercept, Andover, Hampshire. pp. 3-53
12. Ogonnaya, F.C., Abdalla, O., Mujeeb-Kazi, A., Kazi, A.G., Xu, S.S., Gosman, N., Lagudah, E.S., Bonnett, D., Sorrells, M.E. and Tsujimoto, H., 2013. Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.*, 37, pp.35-122.
13. Morgounov, A., Abugalieva, A., Akan, K., Akın, B., Baenziger, S., Bhatta, M., Dababat, A.A., Demir, L., Dutbayev, Y., El Bouhssini, M. and Erginbaş-Orakci, G., 2018. High-yielding winter synthetic hexaploid wheats resistant to multiple diseases and pests. *Plant Genetic Resources*, 16(3), pp.273-278.
14. Mujeeb-Kazi, A., Gul, A., Farooq, M., Rizwan, S. and Ahmad, I., 2008. Rebirth of synthetic hexaploids with global implications for wheat improvement. *Australian Journal of Agricultural Research*, 59(5), pp.391-398.



15. Dreisigacker, S., Kishii, M., Lage, J. and Warburton, M., 2008. Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. *Australian Journal of Agricultural Research*, 59(5), pp.413-420.
16. Lage, J., Warburton, M.L., Crossa, J., Skovmand, B. and Andersen, S.B., 2003. Assessment of genetic diversity in synthetic hexaploid wheats and their *Triticum dicoccum* and *Aegilops tauschii* parents using AFLPs and agronomic traits. *Euphytica*, 134(3), pp.305-317.
17. Zhang, P., Dreisigacker, S., Melchinger, A.E., Reif, J.C., Kazi, A.M., Van Ginkel, M., Hoisington, D. and Warburton, M.L., 2005. Quantifying novel sequence variation and selective advantage in synthetic hexaploid wheats and their backcross-derived lines using SSR markers. *Molecular Breeding*, 15(1), pp.1-10.
18. Cox T., 1997. Deepening the Wheat Gene Pool. *J. Crop Prod*, 1, pp.145–68.
19. Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K., Aktas, H., Ozer, E., Ozdemir, F., Manickavelu, A. and Ban, T., 2015. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of experimental botany*, 66(12), pp.3477-3486.
20. Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C. and Guarino, L., 2017. Past and future use of wild relatives in crop breeding. *Crop Science*, 57(3), pp.1070-1082.
21. Bhatta, M., Morgounov, A., Belamkar, V., Poland, J. and Baenziger, P.S., 2018. Unlocking the novel genetic diversity and population structure of synthetic Hexaploid wheat. *BMC genomics*, 19(1), p.591.
22. Smith, J.B., Schneider, S.H., Oppenheimer, M., Yohe, G.W., Hare, W., Mastrandrea, M.D., Patwardhan, A., Burton, I., Corfee-Morlot, J., Magadza, C.H. and Füssel, H.M.,

2009. Assessing dangerous climate change through an update of the Intergovernmental Panel on Climate Change (IPCC)“reasons for concern”. *Proceedings of the national Academy of Sciences*, 106(11), pp.4133-4137.
23. Zampieri, M., Ceglar, A., Dentener, F. and Toreti, A., 2017. Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environmental Research Letters*, 12(6), p.064008.
24. Becker, S.R., Byrne, P.F., Reid, S.D., Bauerle, W.L., McKay, J.K. and Haley, S.D., 2016. Root traits contributing to drought tolerance of synthetic hexaploid wheat in a greenhouse study. *Euphytica*, 207(1), pp.213-224.
25. Trethowan, R.M. and Mujeeb-Kazi, A., 2008. Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. *Crop Science*, 48(4), pp.1255-1265.
26. Dreccer, M.F., Borgognone, M.G., Ogonnaya, F.C., Trethowan, R.M. and Winter, B., 2007. CIMMYT-selected derived synthetic bread wheats for rainfed environments: yield evaluation in Mexico and Australia. *Field Crops Research*, 100(2-3), pp.218-228.
27. Li, J., WAN, H.S. and YANG, W.Y., 2014. Synthetic hexaploid wheat enhances variation and adaptive evolution of bread wheat in breeding processes. *Journal of Systematics and Evolution*, 52(6), pp.735-742.
28. Das, M.K., Bai, G., Mujeeb-Kazi, A. and Rajaram, S., 2016. Genetic diversity among synthetic hexaploid wheat accessions (*Triticum aestivum*) with resistance to several fungal diseases. *Genetic resources and crop evolution*, 63(8), pp.1285-1296.
29. Ogonnaya, F. C.; Imtiaz, M.; Bariana, H. S.; McLean, M.; Shankar, M. M.; Hollaway, G. J.; Trethowan, R. M.; Lagudah, E. S.; Van Ginkel, M. Mining synthetic hexaploids for

- multiple disease resistance to improve bread wheat. In *Australian Journal of Agricultural Research*; 2008; Vol. 59, pp. 421–431.
30. Jighly, A., Alagu, M., Makdis, F., Singh, M., Singh, S., Emebiri, L.C. and Ogbonnaya, F.C., 2016. Genomic regions conferring resistance to multiple fungal pathogens in synthetic hexaploid wheat. *Molecular breeding*, 36(9), p.127.
31. Zegeye, H., Rasheed, A., Makdis, F., Badebo, A. and Ogbonnaya, F.C., 2014. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One*, 9(8), p.e105593.
32. Villareal, R.L., Mujeeb-Kazi, A., Fuentes-Davila, G., Rajaram, S. and Del Toro, E., 1994. Resistance to karnal bunt (*Tilletia indica* Mitra) in synthetic hexaploid wheats derived from *Triticum turgidum* × *T. tauschii*. *Plant breeding*, 112(1), pp.63-69.
33. White, P.J. and Broadley, M.R., 2009. Biofortification of crops with seven mineral elements often lacking in human diets—iron, zinc, copper, calcium, magnesium, selenium and iodine. *New Phytologist*, 182(1), pp.49-84.
34. Alomari, D.Z., Eggert, K., Von Wirén, N., Pillen, K. and Röder, M.S., 2017. Genome-Wide Association Study of Calcium Accumulation in Grains of European Wheat Cultivars. *Frontiers in plant science*, 8, p.1797.
35. Velu, G., Tutus, Y., Gomez-Becerra, H.F., Hao, Y., Demir, L., Kara, R., Crespo-Herrera, L.A., Orhan, S., Yazici, A., Singh, R.P. and Cakmak, I., 2017. QTL mapping for grain zinc and iron concentrations and zinc efficiency in a tetraploid and hexaploid wheat mapping populations. *Plant and soil*, 411(1-2), pp.81-99.
36. Cakmak, I., 2008. Enrichment of cereal grains with zinc: agronomic or genetic biofortification?. *Plant and soil*, 302(1-2), pp.1-17.

37. Lage, J., Skovmand, B., Peña, R.J. and Andersen, S.B., 2006. Grain quality of emmer wheat derived synthetic hexaploid wheats. *Genetic Resources and Crop Evolution*, 53(5), pp.955-962.
38. Calderini, D.F. and Ortiz-Monasterio, I., 2003. Are synthetic hexaploids a means of increasing grain element concentrations in wheat?. *Euphytica*, 134(2), pp.169-178.

## FIGURES

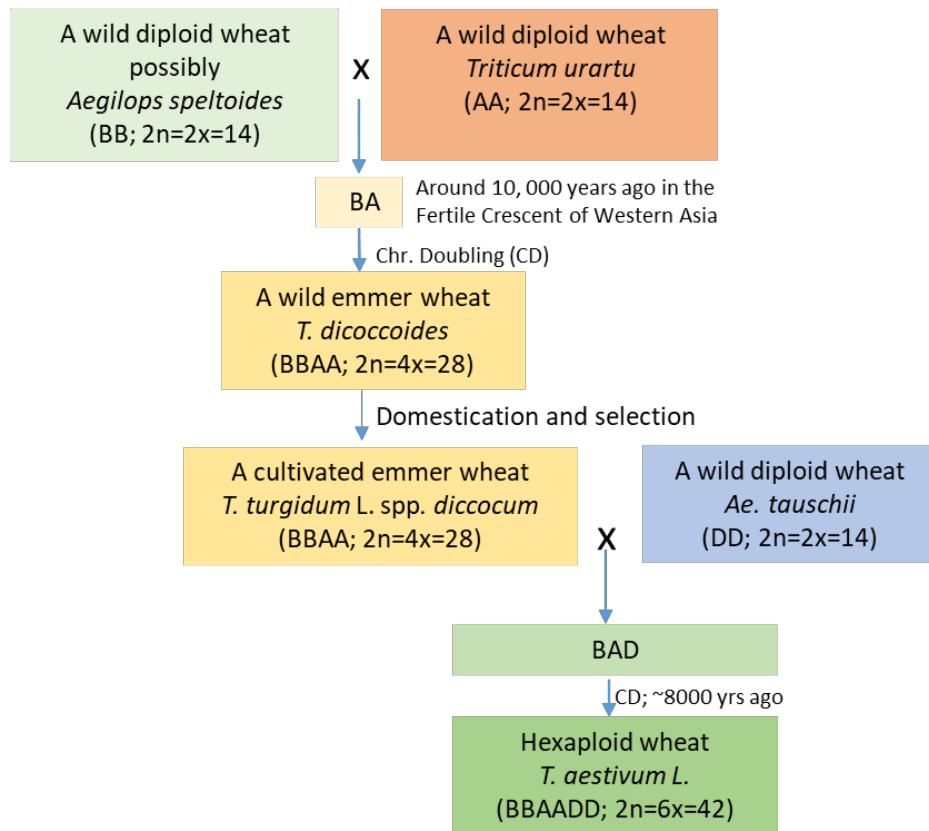


Figure 1. Bread wheat evolution

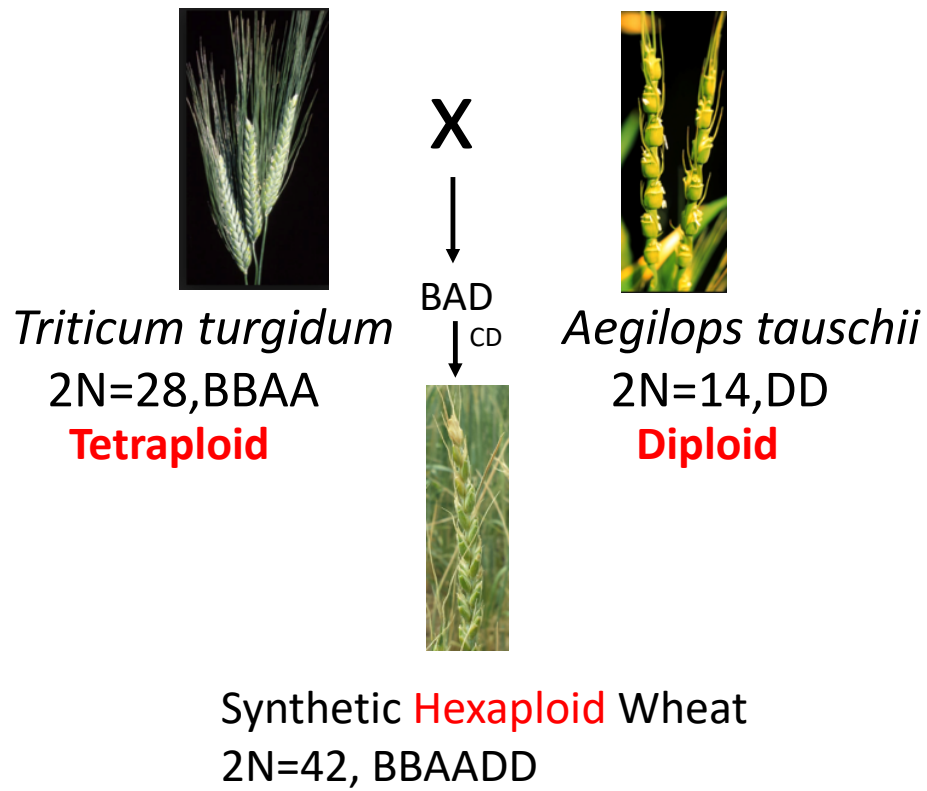


Figure 2. Production of synthetic hexaploid wheat

## CHAPTER 2. UNLOCKING THE NOVEL GENETIC DIVERSITY AND POPULATION STRUCTURE OF SYNTHETIC HEXAPLOID WHEAT

This chapter is published: Bhatta M., Morgounov A., Belamkar V., Poland J., Baenziger P.S., *BMC Genomics* 19:591 (2018). <https://doi.org/10.1186/s12864-018-4969-2>

### ABSTRACT

**Background:** Synthetic hexaploid wheat (SHW) is a reconstitution of hexaploid wheat from its progenitors (*Triticum turgidum* ssp. *durum* L.; AABB x *Aegilops tauschii* Coss.; DD) and has novel sources of genetic diversity for broadening the genetic base of elite bread wheat (BW) germplasm (*T. aestivum* L). Understanding the diversity and population structure of SHWs will facilitate their use in wheat breeding programs. Our objectives were to understand the genetic diversity and population structure of SHWs and compare the genetic diversity of SHWs with elite BW cultivars and demonstrate the potential of SHWs to broaden the genetic base of modern wheat germplasm.

**Results:** The genotyping-by-sequencing of SHW provided 35,939 high-quality single nucleotide polymorphisms (SNPs) that were distributed across the A (33%), B (36%), and D (31%) genomes. The percentage of SNPs on the D genome was nearly same as the other two genomes, unlike in BW cultivars where the D genome polymorphism is generally much lower than the A and B genomes. This indicates the presence of high variation in the D genome in the SHWs. The D genome gene diversity of SHWs was 88.2% higher than that found in a sample of elite BW cultivars. Population structure analysis revealed that SHWs could be separated into two subgroups, mainly differentiated by geographical location of durum parents and growth habit of the crop (spring and winter type). Further population structure analysis of durum and *Ae.* parents separately identified two subgroups, mainly based on type of parents used. Although *Ae. tauschii* parents were divided into two sub-species: *Ae. tauschii* ssp. *tauschii* and ssp. *strangulate*, they were not clearly distinguished in the diversity analysis outcome. Population differentiation between SHWs (Spring\_SHW and Winter\_SHW) samples using analysis of molecular variance indicated 17.43% of genetic variance between populations and the remainder within populations.

**Conclusions:** SHWs were diverse and had a clearly distinguished population structure identified through GBS-derived SNPs. The results of this study will provide valuable information for wheat genetic improvement through inclusion of novel genetic variation and is a prerequisite for association mapping and genomic selection to unravel economically important marker-trait associations and for cultivar development.

**Keywords:** *Aegilops tauschii*, D-genome diversity, genotype-by-sequencing, single nucleotide polymorphism, *Triticum turgidum*, bread wheat

## INTRODUCTION

Hexaploid (bread) wheat (*Triticum aestivum* L.) feeds more than one third of the world's population and is one of the most important staple crops in the world [1]. Bread wheat (BW) evolved from a natural hybridization of the tetraploid cultivated emmer wheat *T. turgidum* L. ssp. *dicoccon* (Schrank) Thell. (2n=28; AABB, a progenitor of modern durum wheat) with the wild diploid *Aegilops tauschii* Coss. (2n=14; DD, goat grass) about 8,000 years ago in the Fertile Crescent [2,3]. Generation of hexaploid wheat from a few accessions of *Ae. tauschii* followed by limited gene flow from *Ae. tauschii* to hexaploid wheat led to limited D-genome diversity [4]. Intercrosses of existing elite wheat germplasm in each breeding cycle and selection has further narrowed the genetic diversity by the depletion of a few alleles from a more diverse gene pool [4]. Such narrow genetic diversity of elite wheat germplasm is a challenge for sustainable wheat production, which is needed for a rapidly growing world population with the predicted dramatic climate changes and other emerging abiotic and biotic stresses.

One approach for broadening the genetic base of BW is utilizing genes from cultivated tetraploid wheat (*T. turgidum*) and from wild relatives (*Ae. tauschii*) through synthetic hexaploid wheat (SHW) production [5–8]. The SHW, often designated as primary synthetic wheat, is a recreation of wheat by crossing between modern durum wheat and wild goat grass. The SHWs provide a rich source of novel



genetic diversity [5–8] and often confer resistance to biotic [9] and abiotic stresses [8,10,11]. The D-genome from SHW is reported to have higher nucleotide sequence diversity than the D-genome from BW [12]. The lack of sequence diversity in the D-genome of BW can be noted from the number of SNPs identified in the A or B genome which usually ranges from two [13,14] to five- [15,16] times higher than SNPs identified in the D genome. Furthermore, *Ae. tauschii* has many desirable genes/alleles for biotic and abiotic stress resistance for wheat improvement [8]. Hence, wheat genome diversity, especially the D-genome diversity, in BW could be improved by crossing to SHW [7]. Introgression of desirable alleles for biotic/abiotic stress resistance and improved end-use quality from wild relatives into elite wheat germplasm is a major objective in many pre-breeding and germplasm development programs [8,10].

Genetic diversity analysis using amplified fragment length polymorphism (AFLP) [5,10] and short sequence repeat (SSR) [5,6,10] have been reported in SHW, however, genetic diversity and population structure analysis of SHWs using single nucleotide polymorphisms (SNPs) are largely unknown. Also, the SHWs used in this study have not been used previously for genetic studies [11]. Therefore, the objectives of this study were to (i) investigate genetic diversity in unique sets of diverse SHW accessions (101) using SNPs derived from genotyping-by-sequencing (GBS) platform, (ii) decipher the presence of population structure in SHW collection, and (iii) compare genetic diversity among SHWs and 12 elite wheat cultivars (comprising 10 cultivars from Lincoln, Nebraska and two from Turkey) to determine the prospects of broadening the genetic base of BW using SHW. Understanding the genetic diversity and population structure of SHWs will help in effectively using these novel genetic resources in breeding programs to broaden the genetic base of wheat, identify novel genes/genomic regions associated with multiple stresses and useful traits, and utilizing such regions/genes in marker assisted breeding.

## METHODS

### Plant Material

Initially, 139 SHWs were analyzed for genetic diversity and population structure (APPENDIX I). However, we found 38 of the entries showed misclassification of durum and *Ae.* parents. Therefore, 38 lines were removed from the analysis and remaining 101 entries were used for the genetic diversity and population structure analysis. Out of 101 SHWs, 15 of them (spring type) originated from one spring durum (Langdon) parent crossed with 15 different *Ae. tauschii* accessions from China, Iran, Kyrgyzstan, Jammu and Kashmir, and Turkmenistan developed by Kyoto University, Japan. The remaining (86) SHW (winter type) originated from the six winter durum parents from Ukraine and Romania (AISBERG, LEUC 84693, PANDUR, UKR-OD 1530.94, UKR-OD 761.93, and UKR-OD 952.92) crossed with 10 different *Ae. tauschii* accessions from Azerbaijan, Iran, Russia, and Unknown; and they were developed by International Maize and Wheat Improvement Center (CIMMYT) from 2004-2013 [11]. Originally, 12 crosses among six durums and 11 *Ae. tauschii* accessions were involved in the creation of 12-winter type SHWs (APPENDIX II). In the early generation of these crosses, due to the segregation, partial sterility and outcrossing, and continuous selection [11], 79 entries were selected as unique lines as they differed phenotypically [11] and on their kinship relationship values. Furthermore, we found seven entries (F8 generation) still segregating (possibly due to outcrossing) for spike color and awn characters in the field experiment conducted in 2016 in Konya, Turkey (APPENDIX I), which were selected as new lines and finally resulted in 86 winter SHWs. The SHWs under study have not been well characterized for genetic studies [11] as might be expected with the continued segregation in the lines. The previously known information of these SHWs were provided in Morgounov et al. [11], who documented that the lines had useful genetic variation for multiple diseases resistance including rust resistance (leaf [incited by *Puccinia triticina*], stripe [incited by *P. striiformis*], and stem rust [incited by *P. graminis*]), common bunt (*Tilletia tritici* and *T. laevis*) resistance, barley yellow dwarf virus resistance and resistance to soil-borne pathogens (cereal cyst nematode [incited by *Heterodera avenae*] and crown rot [incited by *Fusarium*

*pseudograminearum*]), and pest resistance including Hessian fly (*Mayetiola destructor*), sunny pest (*Eurygaster integriceps*), and Russian wheat aphid (*Diuraphi noxia*) resistance. Therefore, these SHWs under study are highly valuable lines for breeding purpose. These SHWs are maintained by the International Winter Wheat Improvement Program (IWWIP) at CIMMYT, Turkey [11]. For genetic diversity comparisons between SHWs and wheat cultivars, 10 elite BW cultivars ('Camelot', 'Cheyenne', 'Freeman', 'Goodstreak', 'Harry', 'Overland', 'Panhandle', 'Robidoux', 'Ruth', and 'Wesley') from Lincoln, Nebraska, USA and two BW cultivars ('Gerek' and 'Karahan') from Turkey were used.

### Genotyping and SNP Discovery

Genomic DNA was extracted from fresh young leaves (approx. 14 days after sowing) using BioSprint<sup>®</sup> 96 Plant Kit (QIAGEN). The GBS libraries were constructed in 96-plex following digestion with the restriction enzymes PstI and MspI [17] at Wheat Genetics Resource Center at Kansas State University (Manhattan, KS). SNP calling was performed using TASSEL v. 5.2.40 GBS v2 Pipeline [18] with physical alignment to wheat reference genome sequence made available by the International Wheat Genome Sequencing Consortium (IWGSC, RefSeq V1.0) in 2017. The SNPs with MAF less than 5% and missing data more than 20% were removed from the analysis. All lines had more than 80 % SNPs called and none were excluded from the analysis. Similarly, for comparing the genetic diversity between SHWs and BW cultivars and analyses specific to the AB or D genomes, GBS derived SNPs were filtered with the same criteria as SHW for genetic diversity analysis.

### Genetic Diversity and Population Structure Analysis

Basic genetic diversity summary statistics including: effective number of alleles, observed heterozygosity, heterozygosity within population (gene diversity), standardized measure of population differentiation ( $F'_{ST}$ ) using AMOVA [19], Nei's standard genetic distance [20], and Jost's index of population differentiation (Jost's D) [21], were calculated for SHWs using GenoDive v 2.0b27 program

[22]. The details of genetic diversity parameters are provided in GenodDive [22]. Average pairwise divergence or observed nucleotide diversity ( $\pi$ ), expected nucleotide diversity or estimated mutation rate ( $\theta$ ) [23], and Tajima's D [24] were calculated in TASSEL v. 5.2.40 [25]. Evolutionary relationship among SHWs were determined by neighbor joining hierarchical cluster analysis based on genetic similarity in TASSEL [25] and a dendrogram was constructed in FigTree V1.4.3 [26]. Analysis of molecular variance was calculated for estimating components of genetic variance among and within population using Arlequin v. 3.5.2.2 [27].

Population structure was inferred using Bayesian clustering algorithm in the program STRUCTURE v 2.3.4 [28] from the command line python program StrAuto [29] and principal coordinate analysis (PCoA) calculated using distance matrix from TASSEL [25]. For identifying the optimal numbers of subpopulations in STRUCTURE and fixation index ( $F_{ST}$ ) of subpopulation, the genotypes were treated as an admixture population with the allele frequencies correlated model with a total of 100,000 burn-in periods followed by 100,000 Markov chain-Monte Carlo iterations for (hypothetical subpopulations)  $K = 1$  to 10 with five independent runs for each  $K$ . The structure output was visualized using StructureHarvester [30] and the number of subpopulations were determined from delta  $K$  model [31]. Kinship relationship matrix was calculated from centered identity by descent method [32] implemented in TASSEL v. 5.2.40 [25].

## RESULTS

To put these results in perspective, there were seven durum wheat parents and 25 different *Ae.* parents for a total of 101 SWHs. Once the cross is made and the chromosomes are doubled, it would be expected that the SWH should be homozygous. However, our phenotypic data and marker data suggested that heterozygous parents, outcrossing, mechanical mixtures, or misclassification occurred (APPENDIX I). This prompted exclusion of 38 lines and the remaining 101 lines were used subsequently in this study.

The GBS derived SNPs were well distributed across the 21 chromosomes in 101 SHWs (Figure 1). The total number of putative SNPs called from 101 SHWs were 129,115. After filtering, 35,939 SNP markers were used for genetic diversity and population structure analysis. The B genome had the highest number of SNPs (12,705, ~36%), followed by the A genome (11,325, ~33%), and the D genome (10,913, ~31%). There were 996 SNPs located in scaffolds that are not anchored to any of the chromosomes. The number of SNPs per chromosome ranged from 733 (4D) to 2,288 (2B) with an average of 1,664 (Figure 1). The ratio of number of B to A genome SNPs was 1.12, the B to D genome was 1.16, and the A to D genome was 1.04. These ratios indicate the number of SNPs on the D genome were nearly equal to SNPs on the A genome and only slightly lower than SNPs on the B genome.

Summary statistics of various genetic diversity estimates for each genome of SHWs had similar values (Table 1). The average effective number of alleles per locus was 1.54. Observed nucleotide diversity or average pairwise divergence ( $\pi$  bp<sup>-1</sup>) and gene diversity ( $H_s$ ) of the SHW genomes were similar and ranged from 0.31 (D genome) to 0.34 (B genome) with an average of 0.33. Expected nucleotide diversity or expected number of polymorphic sites ( $\theta$  bp<sup>-1</sup>) and observed heterozygosity ( $H_o$ ) in SHWs were similar with an average observed heterozygosity of 0.19. Tajima's D ranged from 2.04 (D genome) to 2.40 (B genome) with an average of 2.26. Tajima's D [24] test for selection showed  $D = 2.26$ , that means these genotypes showed significant deviation from the neutral expectation ( $D=0$ ) and rare alleles were present at low frequencies.

### Population Structure

The population structure of 101 SHW was first analyzed on the basis of the ABD genome to study them using all of their genetic diversity. Then the 101 SHWs were analyzed on the basis of the AB and the D genome separately to study genetic diversity of durum and *Ae.* parents, respectively.

*Population structure of the ABD genome (synthetic hexaploid wheat)*

The 101 SHWs showed clear evidence of population structure. Delta K values obtained from the STRUCTURE (Bayesian clustering algorithm) output were used to classify subpopulation. The largest delta K was observed at K=2 (Figure 2A), suggesting the presence of two subpopulations (Figure 2B). The first group contains 15 spring SHWs (syn. SHW developed from Japan), designated as ‘Spring\_SHW’ and second group contains 86 winter SHWs (syn. SHWs developed by CIMMYT), designated as ‘Winter\_SHW’. In Spring\_SHW, all SHWs (15) had the same durum parent ‘Langdon’ developed in North Dakota, USA. In Winter\_SHW, 23 out of 86 SHWs have a durum parent PANDUR developed at Fundulea, Romania and remaining 63 SHWs had durum parents (AISBERG, LECUC.84693, UKR-OD.761.93, UKR-OD.952.92, and UKR-OD.1530.94) developed from Odessa, Ukraine. The growth habit of lines in Spring\_SHW were spring type, whereas lines Winter\_SHW were winter types.

When comparing the grouping obtained from Bayesian clustering in the neighbor joining cluster (Figure 2C) and principal coordinate analysis (PCoA) (Figure 2D), SHWs were again divided into two subgroups (Spring\_SHW and Winter\_SHW) similar to that of Bayesian clustering (Figure 2B).

The population structure of SHWs were mainly grouped based on the geographical location of durum parents and growth habit of the crop. Therefore, the population structure using durum and *Ae. tauschii* were studied separately to further understand how durum or *Aegilops* parents were grouped.

*Population structure using the AB genome (Durum parents)*

When looking at grouping based on the AB genome (durum parent) of SHWs, two groups were obtained from Bayesian clustering (Figures 3A and 3B). The first group contains 15 entries designated as ‘Spring\_Durum’ and second group contains 86 entries, designated as ‘Winter\_Durum’. In Spring\_Durum, all entries (15) have a Langdon durum from North Dakota, USA as a parent. In Winter\_Durum, 23 out of 86 entries have a durum parent from Romania called PANDUR and remaining 63 entries have a parent

from Odessa, Ukraine (AISBERG, LECUC.84693, UKR-OD.761.93, UKR-OD.952.92, and UKR-OD.1530.94). Two subgroups were also classified from the neighbor joining cluster analysis (Figure 3C) and PCoA (Figure 3D), and matched the results obtained from Bayesian clustering algorithm (Figure 2B).

#### *Population structure using the D genome (Aegilops parents)*

When looking at grouping based on the D genome (diploid parent, *Ae. tauschii*) of SHWs, two groups were obtained from Bayesian clustering (Figures 4A and 4B). The first group contains 15 entries designated as ‘Aegilops1’ and second group contains 86 entries, designated as ‘Aegilops2’. In Aegilops1, 8 out of 15 entries were *Ae. tauschii* ssp. *strangulata* and remaining were *Ae. tauschii* ssp. *tauschii* (2) and unknown (5). In Aegilops2, 65 out of 86 entries were *Ae. tauschii* ssp. *tauschii* and remaining were *Ae. tauschii* ssp. *strangulata* (9) and unknown (12). Two subgroups were also classified from the neighbor joining cluster analysis (Figure 4C) and PCoA (Figure 4D), and matched the results obtained from Bayesian clustering algorithm (Figure 2B).

#### Genetic Diversity between the Two Synthetic Hexaploid Wheat Groups

The effective number of alleles across SHW groups ranged from 1.31 to 1.51 (Table 2). Observed heterozygosity ( $H_o$ ) for the two subgroups identified in Bayesian clustering were similar whereas gene diversity ( $H_s$ ) was lower for Spring\_SHW (0.18) group compared to Winter\_SHW (0.31) group. The  $F_{ST}$  (F statistic) obtained from Bayesian clustering measures the divergence and heterogeneity within predefined subgroups and is obtained by estimating the correlation of alleles within the same subgroup [33]. Mean  $F_{ST}$  values were about 66% in Spring\_SHW and 13 % in Winter\_SHW, which indicates population differentiation among genotypes in Winter\_SHW was lower than Spring\_SHW (Table 2). The population differentiation being lower in Winter\_SHW indicates the lines are more similar in this subgroup as compared to lines within Spring\_SHW.

The effective number of alleles across durum subgroups ranged from 1.22 to 1.54 and *Ae.* subgroups ranged from 1.47 to 1.49 (Table 2). Observed heterozygosity within two subgroups of durum and within two subgroups of *Ae.* were similar. Gene diversity of durum subgroups ranged from 0.13 (Spring\_Durum) to 0.32 (Winter\_Durum) and *Ae.* subgroups (Aegilops1 and Aegilops2) was 0.30.

Pairwise population differentiation was obtained from [34] standardized measure of population differentiation ( $F'_{ST}$ ) estimated using an analysis of molecular variance (AMOVA) [19] and this is used for comparison between organisms with different effective population sizes [34], Jost's D [21] as an index of population differentiation that is independent of the amount of within population diversity ( $H_s$ ) computed, and Nei's D [20] as the standard genetic distance was computed from GenoDive (Table 3). Standardized population differentiation ( $F'_{ST}$ ) between Spring\_SHW and Winter\_SHW was 0.34, Spring\_Durum and Winter\_Durum was 0.39, and Aegilops1 and Aegilops2 was 0.22 (Table 3). Similarly, Jost's D (index of population differentiation) between Spring\_SHW and Winter\_SHW was 0.17, Spring\_Durum and Winter\_Durum was 0.20, and Aegilops1 and Aegilops2 was 0.11 (Table 3). The Nei's D (standard genetic distance) between Spring\_SHW and Winter\_SHW was 0.19, Spring\_Durum and Winter\_Durum was 0.22, and Aegilops1 and Aegilops2 was 0.12 (Table 3).

Population differentiation between SHWs (Spring\_SHW and Winter\_SHW) subgroups using analysis of molecular variance (AMOVA) found that 17.43% of the total genetic variance was explained by the differences between subgroups and 82.57% due to the variation within subgroups (Table 4).

### Genetic Diversity of Synthetic Hexaploid Wheat and Bread Wheat Cultivars

For comparing the genetic diversity between SHW and BW, 34,887 high quality SNPs available after quality filtering were used for genetic diversity analysis. The effective number of alleles (Eff-num) was slightly lower for BW (1.26 to 1.38) compared to SHW (1.51 to 1.55) for all genomes (Table 5). The observed heterozygosity of BW ( $H_o$ = 0.17 to 0.18) was slightly lower compared to SHW ( $H_o$ = 0.19 to 0.20) for all genomes. The gene diversity was significantly lower for the BW cultivars ( $H_s$ = 0.17 to 0.25)



compared to SHWs ( $H_s = 0.32$  to  $0.33$ ) for all genomes. Percentage of SHW gene diversity was larger than BW and ranged from 32.0% (B genome) to 88.2% (D genome) higher than that found in BW cultivars. The overall three-genome and the unanchored scaffold (ABD+unmapped) gene diversity of SHW (0.33) was 50.0% larger than that found in the BW cultivars (0.22).

## **DISCUSSION**

### **Population Structure**

The potential use of SHWs in genetic improvement of wheat for biotic and abiotic stresses resistance has been given a priority in many wheat breeding programs [7,8,10,11,35]. This study was designed to provide useful information regarding genetic diversity and population structure of SHWs that could potentially broaden the genetic base of BW germplasm as well help in GWAS to unravel unknown genomic regions or genes associated with economically important multiple traits.

In the present study, ~36,000 GBS derived high quality SNPs obtained from 101 SHWs were used. This study also demonstrates the usefulness of GBS derived SNPs markers for assessing the genetic diversity and population structure. The number of SNPs located on the A, B, and D genome in this study was in agreement with previous studies, where the B genome had the highest number of SNPs, followed by the A and D genome [13,17]. Interestingly, in the present study, the number of SNPs on the D genome was similar to the number of SNPs on the A and B genomes. Generally, in previous studies, the number of SNPs in A or B genome is two [13,14] to five [15,16] times higher than in the D genome. This indicates that the SHWs have higher D genome sequence diversity than other sources. Greater sequence diversity (SNPs) of the D genome in SHWs may support the concept that the D genome has novel genetic variations and desirable genes [8] that can be utilized in elite wheat breeding program for broadening the genetic base. Broadening the genetic base may increase the rate of genetic gain, reduce the D-genome

bottleneck, and help protect wheat from adverse effects of climate change due to currently limited genetic variation for key traits.

Two subgroups obtained from Bayesian clustering algorithm, neighbor joining cluster analysis, and PCoA were mainly divided based on the geographical location of the tetraploid (durum) parents (Romanian and Ukrainian durum in Winter\_SHW group and USA durum in Spring\_SHW group) rather than *Ae.* (diploid) parents. This result agreed with the results of Lage et al. [5], where SHW grouped together based on the geographical origin and presumed similar pedigrees of tetraploid parents. In SHW, two-thirds of the SHWs genome comes from tetraploid wheat (AABB) and one-third of the SHW genome comes from diploid parent (DD). Also, there were fewer durum parents (less diversity compared to *Ae. tauschii*) used in the SHW production which is the likely reason that SHW grouped together based on geographical location of tetraploid parent and growth habit of the crop.

Further population structure analysis was performed using Bayesian clustering algorithm for durum and *Ae.* parents separately to understand how they clustered. Durum parents were divided into two subgroups mainly based on the type/pedigree of durum parents used and its origin. When comparing two subgroups of durum parents to two subgroups of SHWs, all entries of Spring\_Durum were in Spring\_SHW and all entries of Winter\_Durum were in Winter\_SHW. Similarly, the *Ae.* parents also clustered into two subgroups. When comparing two subgroups of Aegilops parents to two subgroups of SHWs, all entries of Aegilops1 were in Spring\_SHW and all entries of Aegilops2 were in Winter\_SHW. Most of the entries of Aegilops1 were *Ae. tauschii ssp. strangulata* and most of the entries of Aegilops2 were *Ae. tauschii ssp. tauschii*. However, there was no distinct clustering based on the area of origin and type of the *Ae. tauschii ssp.* Similar results were obtained in the past [28,30]. For instance, in the study of a diversity panel of 322 *Ae. tauschii*, *Ae. tauschii* were divided into four subgroups and the same *Ae. tauschii ssp.* or from the same area of origin were not clustered together (i.e., *Ae. tauschii ssp. tauschii* and *ssp. strangulata* did not separate entirely into separate clusters) [36]. This result may be potentially attributed to the event of migration leading to a decrease in genetic differentiation among subspecies [37] or wrong

pedigree/classification information at the time of *Ae. tauschii* collections. However, in a different study of 477 *Ae. tauschii* accessions, *Ae. tauschii* were divided into two lineages (*Ae. tauschii ssp. tauschii* and *ssp. strangulata*) having little genetic overlap in the clusters [37].

## Genetic Diversity

Analysis of molecular variation suggested that the population differentiation exists in two subgroups obtained from Bayesian clustering algorithm, where most of the variation was accounted by within population variance. Gene diversity ( $H_s$ ) for each subgroup showed that genetic variation in SHWs ranged from 0.18 (Spring\_SHW) to 0.31 (Winter\_SHW) with an overall gene diversity of 101 SHWs was 0.33. In Spring\_SHW group, only one durum parent was used with different accessions of *Ae. tauschii* parents indicating that genetic variation observed within Spring\_SHW was largely due to *Ae. tauschii* parents (D-genome diversity). Furthermore, the D genome gene diversity within 101 SHWs was 0.31 and genetic diversity of diploid parents (*Ae. tauschii*) from SHWs would expected to be very diverse and novel. The SHWs had a significantly higher level of gene diversity ( $H_s = 0.32$  to  $0.33$ ) compared to elite BW cultivars ( $H_s = 0.17$  to  $0.25$ ) in the present study. Similarly, higher gene diversity in SHWs have been reported in the past using AFLP [5] and SSR markers [6,10,38], indicating the usefulness of SHWs in introducing novel sources of genetic diversity into elite BW germplasm. For instance, gene diversity in 54 SHWs using AFLP marker was 0.39 [5]. Mean genetic diversity in SHWs using SSR markers reported in past were 0.5 [6], 0.61 [38], and 0.69 [10]. In general, the gene diversity of SNP makers is low due to its bi-allelic nature whereas SSR markers are high due to its multi-allelic nature. The gene diversity of SHWs using SNP makers in the past were lower than the present study. For instance, lower genetic diversity in SHWs compared to the present study was reported by Zegeye et al. [9], who evaluated 181 SHWs using 2,590 SNP markers and found the genetic diversity ranged from 0.24 (B genome) to 0.26 (D genome).

The gene diversity of BW cultivars (10 cultivars from Nebraska, USA and two cultivars from Turkey) in our study ranged from 0.17 to 0.25. Similar results were obtained in a study of a diversity

panel of 369 Iranian hexaploid wheat accessions. The gene diversity using SNP markers in Iranian wheat landraces and cultivars ranged from 0.14 to 0.20 [13]. The genetic diversity using GBS derived 20,526 SNPs in 8,416 Mexican wheat landraces ranged from 0.06 to 0.26 [39]. The set of SHWs in our study had greater genetic diversity and was reported to have a multiple resistance to biotic and abiotic stresses [11]. This result suggests that the SHWs under study may provide a novel source of genetic diversity (novel alleles for a trait of interest) to the elite wheat breeding program.

## CONCLUSIONS

The present study provided a detailed understanding of genetic diversity and population structure of 101 SHWs and revealed high genetic diversity in the SHW compared to elite BW cultivars. Population structure analysis revealed that SHWs developed from diverse accessions of durum wheat and *Ae. tauschii* originated from different countries were divided into two (Spring\_SHW and Winter\_SHW) distinct groups based mainly on geographical location of durum parents and growth habit of the crop. Further population structure analysis of durum and *Ae.* parents separately identified two subgroups, mainly based on type/pedigree or origin of parents used. Although *Ae. tauschii* parents were divided into two groups mainly based on type of parent used, *Ae. tauschii ssp. tauschii* and *ssp. strangulata* did not separate entirely in each subgroup. The GBS derived SNPs were able to identify the inaccurate pedigree of synthetic hexaploid wheats based on misclassifications of some of the durum or *Ae.* parents from analyzing 139 SHWs and such misclassifications may have resulted due to heterozygous or heterogeneous parent lines, partial sterility and outcrossing, or seed handling error/mechanical mixing. This study found that the percentage of SNPs on the D genome was nearly equal to that of other two genomes (A and B), which is unique to the SHWs as compared to previous studies on BW that reported that the D genome had less than 50% of the SNPs compared to A and B genomes. This result indicated the presence of high variation in the D genome in the SHWs. Furthermore, the gene diversity of SHWs

under study was higher in all three genomes compared to elite BW cultivars and the greatest increase in gene diversity (88.2%) was observed in the D genome of SHWs compared to BW cultivars. Such higher genetic diversity in SHWs suggests that the diversity could be utilized in the elite wheat-breeding program to broaden the genetic diversity and increase genetic gain. The markers with high genome coverage such as GBS derived SNP markers are helpful for elucidating the population structure and genetic diversity. The results of this study will provide valuable information for wheat genetic improvement through inclusion of novel genetic variation and facilitate the discovery of novel source of genes/genomic regions conferring resistance to multiple biotic and abiotic stresses from association mapping study to unravel economically important marker-trait associations.

## REFERENCES

1. Bhatta M, Regassa T, Rose DJ, Baenziger PS, Eskridge KM, Santra DK, et al. Genotype, environment, seeding rate, and top-dressed nitrogen effects on end-use quality of modern Nebraska winter wheat. *J. Sci. Food Agric.* 2017;97:5311–8.
2. Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W. Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* 2002;3:429–41.
3. Dubcovsky J, Dvorak J. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science.* 2007;316:1862–6.
4. Cox T. Deepening the wheat gene pool. *J. Crop Prod.* 1997;1:145–68.
5. Lage J, Warburton ML, Crossa J, Skovmand B, Andersen SB. Assessment of genetic diversity in synthetic hexaploid wheats and their *Triticum dicoccum* and *Aegilops tauschii* parents using AFLPs and agronomic traits. *Euphytica.* 2003;134:305–17.
6. Zhang P, Dreisigacker S, Melchinger AE, Reif JC, Mujeeb Kazi A, Van Ginkel M, et al. Quantifying novel sequence variation and selective advantage in synthetic hexaploid wheats and their backcross-derived lines using SSR markers. *Mol. Breed.* 2005;15:1–10.
7. Dreisigacker S, Kishii M, Lage J, Warburton M. Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. *Aust. J. Agric. Res.* 2008. p. 413–20.
8. Ogonnaya FC, Abdalla O, Mujeeb-Kazi A, Kazi AG, Xu SS, Gosman N, et al. Synthetic hexaploids: Harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.* 2013;37:35–122.
9. Zegeye H, Rasheed A, Makdis F, Badebo A, Ogonnaya FC. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One.* 2014;9.

10. Das MK, Bai G, Mujeeb-Kazi A, Rajaram S. Genetic diversity among synthetic hexaploid wheat accessions (*Triticum aestivum*) with resistance to several fungal diseases. *Genet. Resour. Crop Evol.* 2016;63:1285–96.
11. Morgounov A, Abugalieva A, Akan K, Akın B, Baenziger S, Bhatta M, et al. High-yielding winter synthetic hexaploid wheats resistant to multiple diseases and pests. *Plant Genet. Resour.* 2018;16:273–8.
12. Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, et al. Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. *Genetics.* 2004;167:941–7.
13. 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. *Front. Plant Sci.* 2017;8.
14. Iehisa JCM, Shimizu A, Sato K, Nishijima R, Sakaguchi K, Matsuda R, et al. Genome-wide marker development for the wheat D genome based on single nucleotide polymorphisms identified from transcripts in the wild wheat progenitor *Aegilops tauschii*. *Theor. Appl. Genet.* 2014;127:261–71.
15. Allen AM, Barker GLA, Wilkinson P, Burrige A, Winfield M, Coghill J, et al. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 2013;11:279–95.
16. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci.* 2013;110:8057–62.
17. Poland JA, Brown PJ, Sorrells ME, Jannink J. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One.* 2012;7.
18. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One.* 2014;9. e90346.

19. Meirmans PG. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*. 2006;60:2399–402.
20. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978. 583–90.
21. Jost L. GST and its relatives do not measure differentiation. *Mol. Ecol.* 2008;17:4015–26.
22. Meirmans PG, Van Tienderen PH. GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes*. 2004;4:792–4.
23. Kimura M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*. 1969;61:893–903.
24. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
25. 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;23:2633–5.
26. Rambaut A. FigTree v1.4.3. *Mol. Evol. phylogenetics Epidemiol.* 2016. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>
27. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 2010;10:564–7.
28. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
29. Chhatre VE, Emerson KJ. StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*. 2017;18.



30. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 2012;4:359–61.
31. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 2005;14:2611–20.
32. Endelman JB, Jannink J-L. Shrinkage estimation of the realized relationship matrix. *G3:Genes, Genomes, Genetics.* 2012;2:1405–13.
33. Wright S. The genetical structure of natural populations. *Ann. Eugen.* 1951;15:323–354.
34. Hedrick PW. A standardized genetic differentiation measure. *Evolution.* 2005;59:1633–8.
35. Warburton ML, Crossa J, Franco J, Kazi M, Trethowan R, Rajaram S, et al. Bringing wild relatives back into the family: Recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica.* 2006;149:289–301.
36. Liu Y, Wang L, Mao S, Liu K, Lu Y, Wang J, et al. Genome-wide association study of 29 morphological traits in *Aegilops tauschii*. *Sci. Rep.* 2015;5:15562.
37. Wang J, Luo MC, Chen Z, You FM, Wei Y, Zheng Y, et al. *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* 2013;198:925–37.
38. Hanif U, Rasheed A, Kazi AG, Afzal F, Khalid M, Munir M, et al. Analysis of genetic diversity in synthetic wheat assemblage (*T. turgidum* x *Aegilops tauschii*;  $2n=6x=42$ ; AABBDD) for winter wheat breeding. *Cytologia.* 2014;79:485–500.
39. Vikram P, Franco J, Burgueño-Ferreira J, Li H, Sehgal D, Saint Pierre C, et al. Unlocking the genetic diversity of Creole wheats. *Sci. Rep.* 2016;6:23092.

## TABLES

Table 1. Distribution of SNP markers and genetic diversity summary statistics of 101 synthetic hexaploid wheats including observed nucleotide diversity ( $\pi$  bp<sup>-1</sup>), expected nucleotide diversity ( $\theta$  bp<sup>-1</sup>), Tajima's D, effective number of alleles (Eff-num), observed heterozygosity ( $H_o$ ), and gene diversity ( $H_s$ ).

Genome	No. of SNPs	$\pi$ bp <sup>-1</sup>	$\theta$ bp <sup>-1</sup>	Tajima's D	Eff_Num	$H_o$	$H_s$
A	11325	0.33	0.20	2.34	1.55	0.19	0.33
B	12705	0.34	0.20	2.40	1.56	0.18	0.34
D	10913	0.32	0.20	2.04	1.51	0.19	0.31
AB	24030	0.33	0.20	2.38	1.55	0.19	0.33
ABD+Unmapped	35939	0.33	0.20	2.27	1.54	0.19	0.33

Table 2. Population genetic diversity summary statistics of two subgroups in 101 synthetic hexaploid wheats (SHWs), Durum wheat, and *Aegilops tauschii* (*Aegilops*) obtained from GenoDive including effective number of alleles (Eff-num), observed heterozygosity ( $H_o$ ), and gene diversity ( $H_s$ ), and Mean  $F_{ST}$  obtained from STRUCTURE.

Group	Subgroup	Number of genotypes	Eff_num	$H_o$	$H_s$	Mean $F_{ST}$
Synthetic Hexaploid Wheat (ABD)						
	Spring_SHW	15	1.31	0.18	0.18	0.66
	Winter_SHW	86	1.51	0.19	0.31	0.13
Durum wheat (AB)						
	Spring_Durum	15	1.22	0.17	0.13	0.79
	Winter_Durum	86	1.54	0.19	0.32	0.10
<i>Aegilops tauschii</i> (D)						
	Aegilops1	15	1.49	0.20	0.30	0.30
	Aegilops2	86	1.47	0.19	0.30	0.27

Table 3. A standardized measure of population differentiation ( $F'_{ST}$ ), Jost's D as an index of population differentiation, and Nei's D as the standard genetic distance in two subgroups in 101 synthetic hexaploid wheats (SHWs) was computed in GenoDive.

Population	$F'_{ST}$	Jost's D	Nei's D
Spring_SHW and Winter_SHW	0.34	0.17	0.19
Spring_Durum and Winter_Durum	0.39	0.20	0.22
Aegilops1 and Aegilops2	0.22	0.11	0.12

Table 4. Analysis of molecular variance (AMOVA) within and between the two subgroups of 101 synthetic hexaploid wheats identified by the Bayesian clustering.

Source	d.f.	Sum of squares	Mean squares	Estimated variation	Percentage of variation (%)
Between Populations	1	60448.12	60448.12	560.03	17.43
Within populations	99	319345.46	3225.71	2652.08	82.57
Total	100	379793.58	-	3212.11	

Table 5. Population genetic diversity summary statistics of 101 synthetic hexaploid wheats (SHWs) and 12 bread wheat cultivars obtained from GenoDive.

Genome	Population	No. of Genotypes	SNPs used	Eff_num <sup>a</sup>	Ho <sup>b</sup>	Hs <sup>c</sup>	Gene diversity of SHW increased compared to BW (%)
A			11297				
	Synthetic Hexaploid wheat	101		1.54	0.19	0.33	37.5
	Bread Wheat	12		1.38	0.17	0.24	
B			11297				
	Synthetic Hexaploid wheat	101		1.54	0.19	0.33	32.0
	Bread Wheat	12		1.38	0.18	0.25	
D			10008				
	Synthetic Hexaploid wheat	101		1.51	0.20	0.32	88.2
	Bread Wheat	12		1.26	0.17	0.17	
AB			23930				
	Synthetic Hexaploid wheat	101		1.55	0.19	0.33	32.0
	Bread Wheat	12		1.38	0.17	0.25	
ABD			33938				
	Synthetic Hexaploid wheat	101		1.54	0.19	0.33	50.0
	Bread Wheat	12		1.35	0.17	0.22	
ABD+unmapped			34887				
	Synthetic Hexaploid wheat	101		1.54	0.19	0.33	50.0
	Bread Wheat	12		1.35	0.17	0.22	

<sup>a</sup>Eff-num: effective number of alleles, <sup>b</sup>Ho: observed heterozygosity, and <sup>c</sup>Hs: gene diversity

## FIGURES

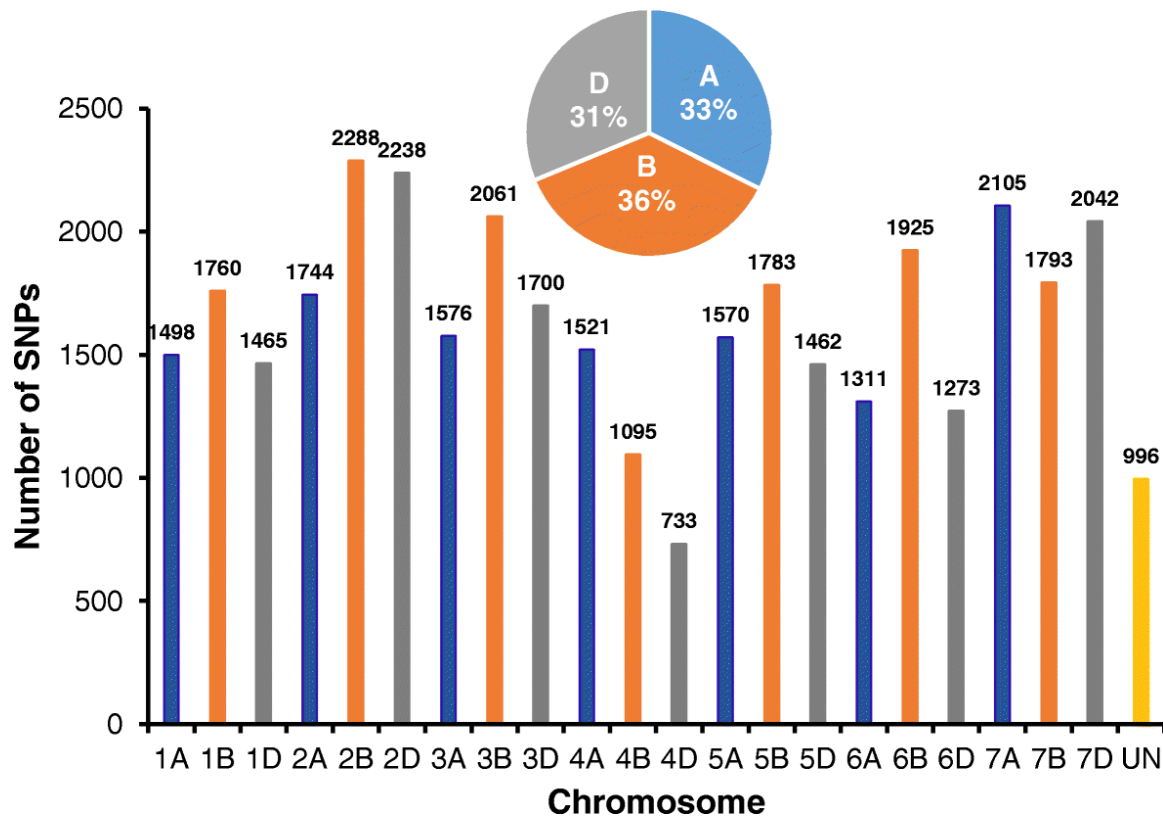


Figure 1. Distribution of 35,939 single nucleotide polymorphisms (SNPs) across 21 chromosomes and unanchored scaffolds in 101 synthetic hexaploid wheats.

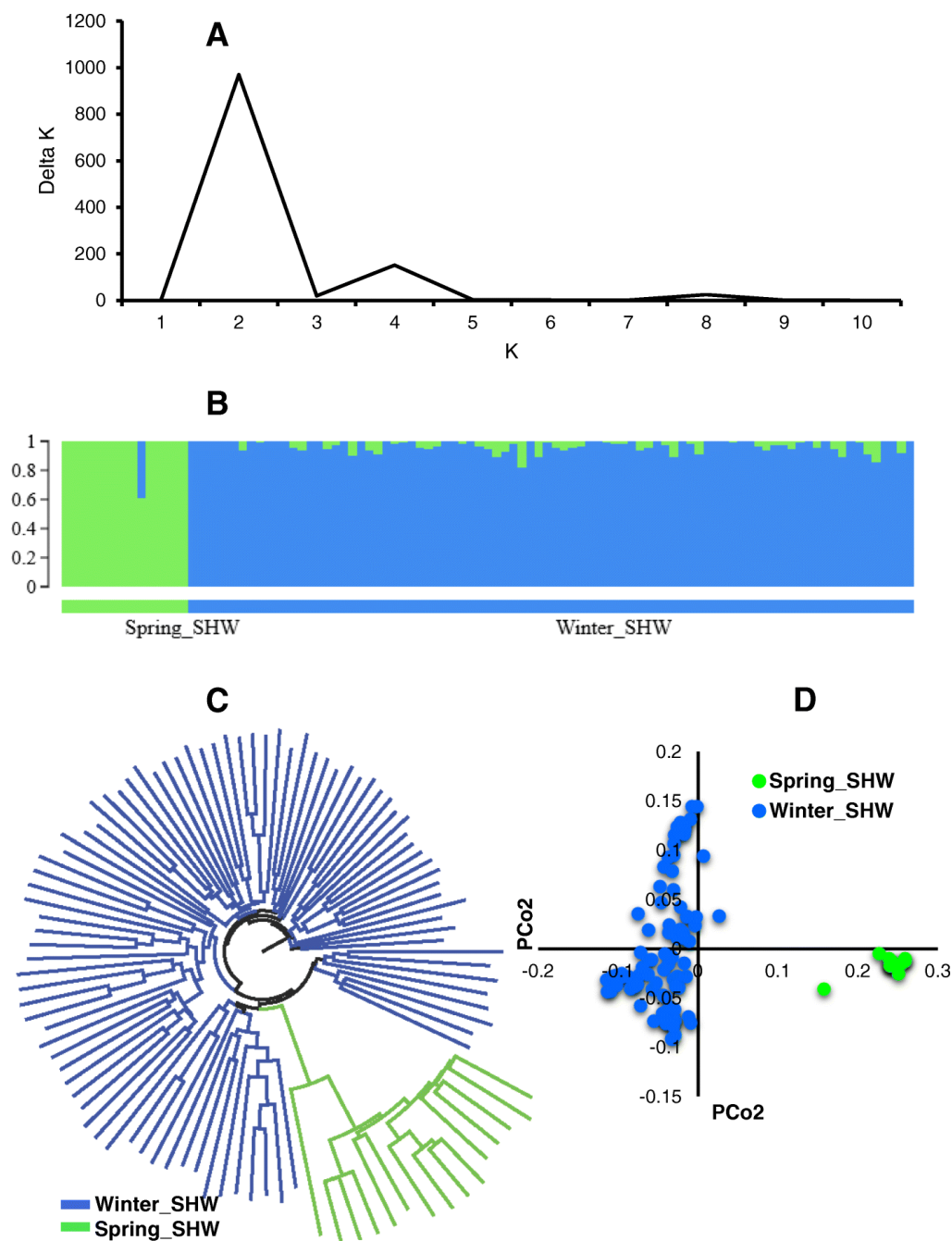


Figure 2. Population structure of the 101 synthetic hexaploid wheat germplasm. A: Line graph of  $\Delta K$  over  $K$  from 1 to 10, and the highest peak was observed at  $\Delta K=2$ , suggesting the synthetic hexaploid wheat (SHW) germplasm has two subgroups. B: The two subgroups identified from the STRUCTURE and grouped based on the geographical location of the durum parents and growth habit of the crop. C: Cluster analysis (neighbor joining) and D: Principal coordinate analysis (PCoA). Color reflects grouping derived from STRUCTURE.



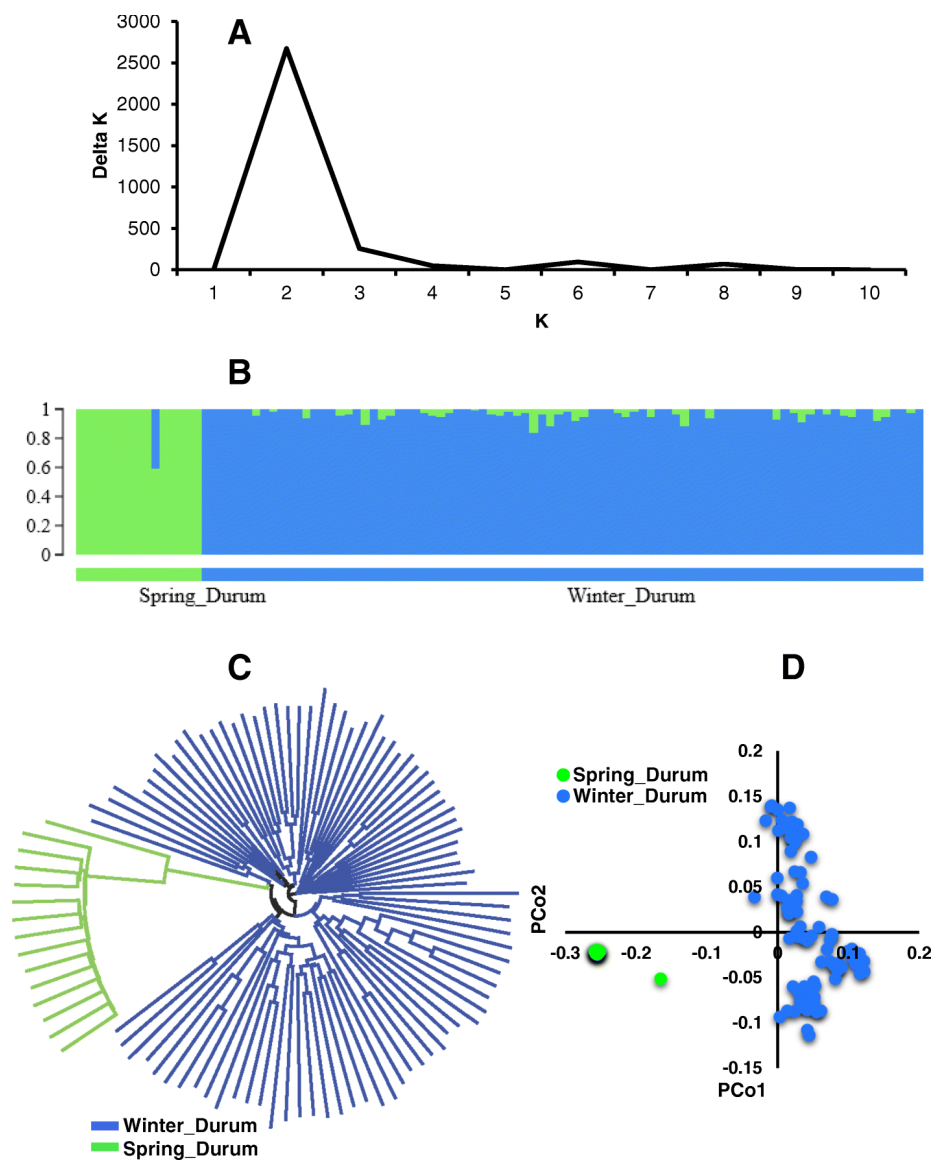


Figure 3. Population structure of the durum parents used in the production of synthetic hexaploid wheat (SHW) germplasm. A: Line graph of delta K over K from 1 to 10, and the highest peak was observed at Delta K=2, suggesting the durum wheat used in this study has two subgroups. B: The two subgroups were identified from the STRUCTURE and grouped based on the geographical location of the durum parents and growth habit of the crop. C: Cluster analysis (neighbor joining) and D: Principal coordinate analysis (PCoA). Color reflects grouping derived from STRUCTURE.

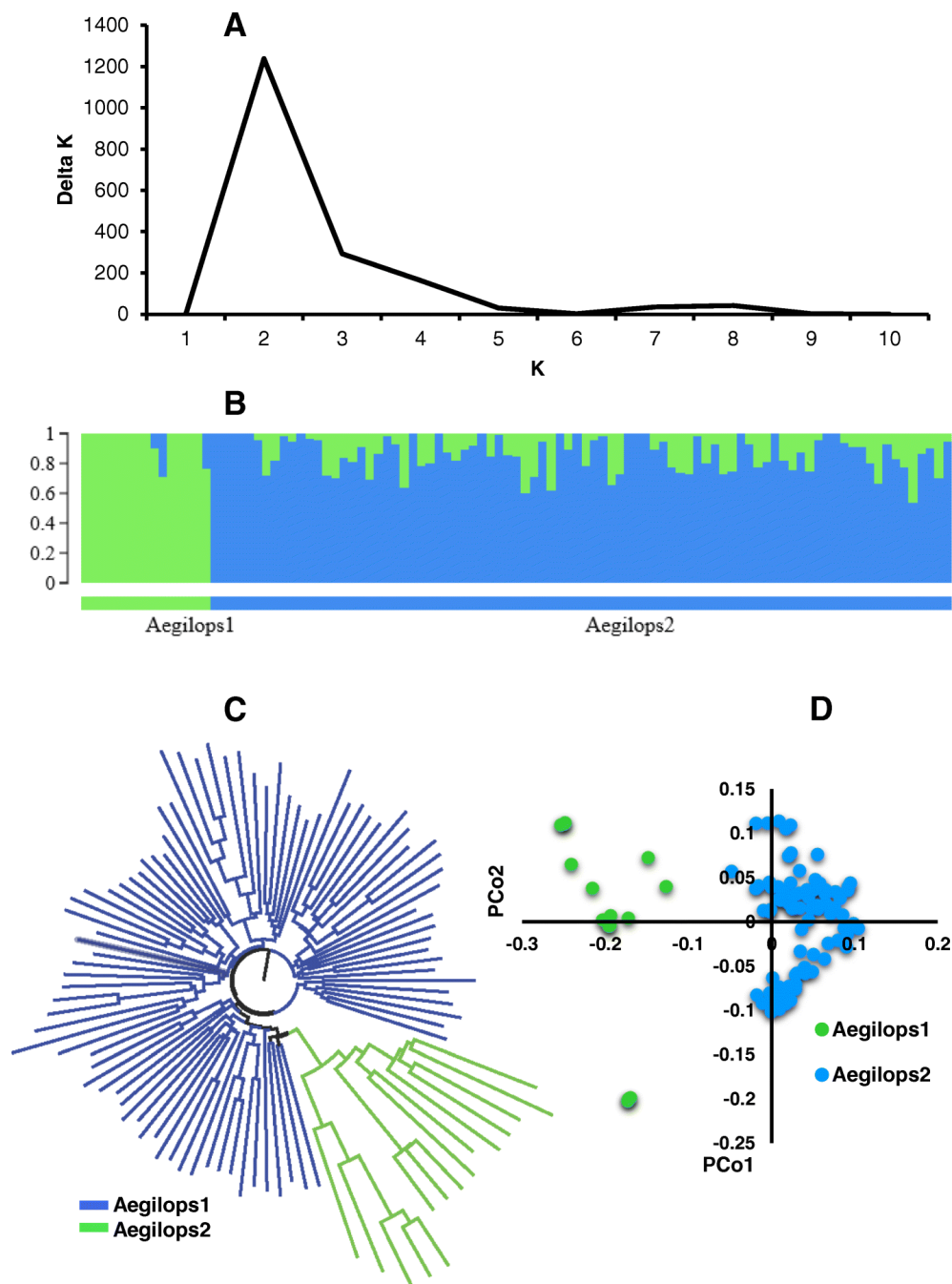


Figure 4. Population structure of the *Aegilops* parents used in the production of synthetic hexaploid wheat. A: Line graph of delta K over K from 1 to 10, and the highest peak was observed at Delta K=2, suggesting the *Aegilops* used in this study has two subgroups. B: The two subgroups were identified from the STRUCTURE and grouped based on the type of *Aegilops* parents used. C: Cluster analysis (neighbor joining) and D: Principal coordinate analysis (PCoA). Color reflects grouping derived from STRUCTURE.

### CHAPTER 3. GENOME-WIDE ASSOCIATION STUDY REVEALS NOVEL GENOMIC REGIONS FOR GRAIN YIELD AND YIELD-RELATED TRAITS IN DROUGHT-STRESSED SYNTHETIC HEXAPLOID WHEAT

This chapter is published: Bhatta M., Morgounov A., Belamkar V., Baenziger P.S. *International Journal of Molecular Sciences* 19:3011 (2018). <https://doi.org/10.3390/ijms19103011>

#### ABSTRACT

Synthetic hexaploid wheat (SHW;  $2n=6x=42$ , AABBDD, *Triticum aestivum* L.) is produced from an interspecific cross between durum wheat ( $2n=4x=28$ , AABB, *T. turgidum* L.) and goat grass ( $2n=2x=14$ , DD, *Aegilops tauschii* Coss.) and is reported to have significant novel alleles controlling biotic and abiotic stresses resistance. A genome-wide association study (GWAS) was conducted to unravel these loci [marker-trait associations: MTAs] using 35,648 genotyping-by-sequencing derived single nucleotide polymorphisms in 123 SHWs. We identified 90 novel MTAs (45, 11, and 34 on the A, B, and D genome, respectively) and haplotype blocks associated with grain yield and yield-related traits including root-traits under drought stress. The phenotypic variance explained by the MTAs ranged from 1.1 to 32.3%. Most of the MTAs (120 out of 194) identified were found within genes, and of these 45 MTAs were in genes annotated as having a potential role in drought stress. This result provides further evidence for the reliability of MTAs identified. The large number of MTAs (53) identified especially on the D genome demonstrate the potential of SHWs for elucidating the genetic architecture of complex traits and provide an opportunity for further improvement of wheat under rapidly changing climatic conditions.

**Keywords:** marker-trait association; haplotype block; genes; root traits; D-genome; GBS; single nucleotide polymorphisms; durum wheat; bread wheat; complex traits

## Introduction

Drought is one of the most important abiotic stress that reduces crop productivity and is expected to increase with the change in climate [1]. Erratic rainfall patterns caused by climate change may aggravate drought stress and that will have a major impact on agriculture [2,3]. The most prominent example of the impact of drought stress on agriculture was the 2012 drought stress in the United States, where moderate to extreme drought stress occurred across the central agricultural states that resulted in crop harvest failure for corn (*Zea mays* L.), sorghum (*Sorghum bicolor* L.), and soybean (*Glycine max* L.), and the agriculture loss due to drought was estimated to be \$30 billion [4]. To cope up with the challenges of drought stress, plant breeders have been focusing on improving drought tolerance since several decades [2,3,5]. However, the drought tolerance is a complex phenomenon as most of the traits associated with drought tolerance are polygenic in nature and understating the genetic architecture of drought tolerance is still underway [3] including in wheat (*Triticum sps.*) [2]. Wheat is one of the most important staple cereal crops mainly grown under rainfed conditions [3,6] and is expected to suffer from drought stress [3]. Therefore, breeding for drought tolerance and identifying genomic regions and underlying candidate genes associated with drought tolerance is important for wheat improvement.

Bread wheat (*T. aestivum* L.) has limited genetic and phenotypic diversity available for breeding for drought tolerance [2]. This is mainly due to the genetic bottleneck experienced during its origin and subsequent domestication [7,8]. Diversity can be increased through the production of synthetic hexaploid wheat (SHW) and its utilization in breeding programs [9–11]. Synthetic hexaploid wheat ( $2n=6x=42$ , AABBDD) is produced from an interspecific cross between durum wheat ( $2n=4x=28$ , AABB, *T. turgidum* L.) and goat grass ( $2n=2x=14$ , DD,

*Aegilops tauschii* Coss.). The SHWs are reported to have significant genetic variation for biotic [12,13] and abiotic stress resistance [10,14,15]. However, previous studies mostly focused mainly on biotic stresses including leaf rust (incited by *Puccinia triticina*) [13,16,17], stem rust (incited by *P. graminis*) [16,17], stripe rust (incited by *P. striiformis*) [12,16,17], Fusarium head blight (incited by *Fusarium graminearum*) [13], yellow spot (incited by *Pyrenophora tritici-repentis*) [16,17], Septoria nodorum (incited by *Parastagonospora nodorum*) [16,17], Septoria tritici blotch (incited by *Mycosphaerella graminicola*) [13,16], cereal cyst nematode (incited by *Heterodera avenae*) [16], crown rot (incited by *F. pseudograminearum*) [17], and root-lesion nematode (incited by *Pratylenchus thornei* and *P. neglectus*) [16]. Therefore, exploiting genetic variation under abiotic stress such as drought is timely and needed to further utilize the potential of SHWs.

About 800 quantitative trait loci (QTLs) and marker trait associations (MTAs) have been reported for drought tolerant traits (agronomic, physiological, root, and yield-related traits) using bi-parental mapping (~691 QTLs) and genome wide association studies (GWAS; ~109 MTAs) in wheat [3]. Only 68 QTLs are major QTLs that explain more than 19% of phenotypic variation [3]. This study was conducted to identify novel genomic regions associated with grain yield and yield-related traits using GWAS performed using 35,648 GBS-derived SNPs in 123 SHWs grown under two drought stressed growing seasons (2016 and 2017) in Konya, Turkey. Subsequently, the underlying genes for the MTAs identified were investigated for their potential role in drought stress using the functional annotations. To the best of our knowledge, this is the first report on GWAS on grain yield and yield-related traits under drought stress in SHWs. The results from this study will be a valuable resource for the genetic improvement of grain yield and

yield-related traits in drought stress, introgression of desirable genes from SHWs into elite wheat germplasm, genomic selection, and marker-assisted selection in the breeding program.

## Materials and Methods

### Site Description

A field experiment was conducted during two growing seasons (2016 and 2017) under drought stressed conditions (rainfed) at the research farm located at Bahri Dagdas International Agricultural Research Institute in Konya, Turkey (37°51'15.894" N, 32°34'3.936" E; Elevation = 1,021 m). This site was characterized by a low precipitation (below 300 mm), low humidity, and slightly alkaline clay loam soil [18].

### Plant Materials and Experimental Design

One hundred twenty-three SHWs developed from two introgression programs were used (APPENDIX I). The first group was developed by Kyoto University, Japan from one spring durum ('Langdon') parent crossed with 14 different *Ae. tauschii* accessions resulting in 14 different lines (Table S2). The remaining 109 lines were the second group of synthetics that were developed by CIMMYT from the six-durum parents crossed with 11 different *Ae. tauschii* accessions mainly from the Caspian Sea basin area. Initially, 13 crosses among six winter durum wheats were involved in the creation of 13 different winter type synthetics. However, due to the partial sterility, segregation, and continuous selection in the early generation, 109 lines were selected as unique lines because of their difference in phenotype [14] and their kinship values [11]. The synthetic genotypes used in this study are unique and have been developed recently

and tested for their agronomic traits [14], genetic diversity, and population structure [11]. The detailed information of these SHWs were provided in Bhatta et al. [11].

The experimental design was an augmented design (plot size: 1.2 m x 5 m) with replicated checks ('Gerek' and 'Karahana') in the 2016 growing season and modified alpha-lattice design (plot size: 1.2 m x 5 m) including 123 SHWs and replicated checks (Gerek and Karahana) with two replications in the 2017 growing season. SHWs were planted on 20 September in 2015 and harvested on 18 July 2016 for the 2016 growing season, whereas SHWs were planted on 15 September in 2016 and harvested on 21 July 2017 for the 2017 growing season.

#### Trait Measurements

Grain yield (GY) was obtained by harvesting four middle rows of 1.008 m<sup>2</sup> (i.e., 84 x 120) and reported in g m<sup>-2</sup>. Harvest index (HI), dry biomass weight (BMWT), thousand kernel weight (TKW), grain volume weight (GVWT), awn length (AWNLN), Flag leaf length (FLLN), flag leaf width (FLW), flag leaf area (FLA=0.8 x FLLN x FLW) were measured using previously reported protocols [19–22]. Stem diameter (STMDIA) was measured from five randomly selected plants per plot using digital Vernier caliper at second internode from the soil surface at physiological maturity. Root length (RTLN) was measured from randomly selected three plants per plot after 3-4 days of flowering (Zadoks 60 growth stage) using WinRhizo software (WinRhizo reg. 2009c, Regent Instruments Inc., Quebec City, Quebec, Canada).

#### Phenotypic Data Analysis

Combined analysis of variance (ANOVA) was performed using the following model:

$$y_{ijklmn} = \mu + Yr_i + R(Yr)_{ji} + B(R(Yr))_{kji} + C_1 + G_{m(kji)} + GXYr_{mi} + e_{ijklmn} \quad 1)$$

Where,  $y_{ijklmn}$  is the grain yield-related trait;  $\mu$  is overall mean;  $Yr_i$  is the effect of  $i^{\text{th}}$  year;  $R(Yr)_{ji}$  is the effect of  $j^{\text{th}}$  replication within  $i^{\text{th}}$  year;  $B(R(Yr))_{kji}$  is the effect of  $k^{\text{th}}$  incomplete block within  $j^{\text{th}}$  replication of  $i^{\text{th}}$  environment;  $C_1$  is the  $1^{\text{th}}$  checks;  $G_{m(kji)}$  is the effect of  $m^{\text{th}}$  genotypes (new variable, where check is coded as 0 and entry is coded as 1 and genotype was taken as new variable x entry) within  $k^{\text{th}}$  incomplete block of  $j^{\text{th}}$  replication in  $i^{\text{th}}$  year;  $GxYr_{mi}$  is the interaction effect of  $m^{\text{th}}$  genotype and  $i^{\text{th}}$  year;  $e_{ijklmn}$  is the residual. In combined ANOVA, year and check were assumed as fixed effects whereas genotype, genotype x year interaction, replication nested within a year, and incomplete block nested within replications were assumed as random effects.

Individual analyses of variance were performed because most of the traits had highly significant genotype by year interaction and therefore will be discussed hereafter. In the year 2016, an augmented design was analyzed using the following model for the estimation of best linear unbiased predictors (BLUPs):

$$y_{ijkl} = \mu + B_i + C_j + G_{k(i)} + e_{ijkl} \quad 2)$$

Where  $y_{ijkl}$  is the trait,  $\mu$  is overall mean;  $B_i$  is the effect of  $i^{\text{th}}$  incomplete block;  $C_j$  is the  $j^{\text{th}}$  checks;  $G_{k(i)}$  (new variable, where check is coded as 0 and entry is coded as 1 and genotype was taken as new variable x entry) is the effect of  $k^{\text{th}}$  genotypes within  $i^{\text{th}}$  block;  $e_{ijkl}$  is the residual. In ANOVA calculated for the 2016 datasets, the check was assumed as fixed effect whereas genotype and incomplete block were assumed as random effects.



In the year 2017, alpha ( $\alpha$ ) lattice design with two replications was analyzed using the following model for the estimation of BLUPs:

$$y_{ijklm} = \mu + R_i + B(R)_{ji} + C_k + G_{l(ji)} + e_{ijklm} \quad 3)$$

Where,  $y_{ijklm}$  is the trait,  $\mu$  is overall mean;  $R_i$  is the effect of  $i^{\text{th}}$  replication;  $B(R)_{ji}$  is the effect of  $j^{\text{th}}$  block within  $i^{\text{th}}$  replication;  $C_k$  is the  $k^{\text{th}}$  checks;  $G_{l(ji)}$  (new variable, where check is coded as 0 and entry is coded as 1 and genotype was taken as new variable x entry) is the effect of  $k^{\text{th}}$  genotypes within  $j^{\text{th}}$  incomplete block of  $i^{\text{th}}$  replication;  $e_{ijklm}$  is the residual. In ANOVA calculated for the 2017 datasets, the check was assumed as fixed effect whereas genotype, replication, and incomplete block nested within replication were assumed as random effects.

All phenotypic data were analyzed using PROC MIXED in SAS 9.4 (SAS Institute Inc., Cary, NC) [23] using restricted maximum likelihood (REML) approach unless mentioned otherwise.

Broad-sense heritability for each trait in each year was calculated based on entry mean basis using the following formula for 2016, 2017, and combined experiments, respectively:

$$H^2 = \frac{\sigma^2 g}{\sigma^2 g + \sigma^2 e} \quad 4)$$

$$H^2 = \frac{\sigma^2 g}{\sigma^2 g + \frac{\sigma^2 e}{r}} \quad 5)$$

$$H^2 = \frac{\sigma^2 g}{\sigma^2 g + \frac{\sigma^2 gxyr}{n} + \frac{\sigma^2 e}{nr}} \quad 6)$$

Where,  $\sigma^2_g$ ,  $\sigma^2_{g \times yr}$ , and  $\sigma^2_e$  are the variance components for genotype, genotype x year, and error, respectively, whereas n and r are the number of years and replications, respectively.

Pearson's correlation of grain yield and related traits was calculated based on BLUPs for each trait in each year using PROC CORR in SAS. The principal component biplot analysis (PCA-biplot) was performed based on the correlation matrix to avoid any variation due to the different scales of the measured variables using 'factoextra' package in R software [24].

### Genotyping and SNP Discovery

Genomic DNA was extracted from two to three fresh young leaves of 14 day old seedlings using BioSprint 96 Plant Kits (Qiagen, Hombrechtikon, Switzerland), as described in Bhatta et al. [11]. The GBS libraries were constructed in 96-plex following digestion with two restriction enzymes, PstI and MspI [25] and pooled libraries were sequenced using Illumina, Inc. (San Diego, CA) next generation sequencing platforms at the Wheat Genetics Resource Center at Kansas State University (Manhattan, KS). SNP call was performed using TASSEL v. 5.2.40 GBS v2 Pipeline [26] with physical alignment to *Chinese Spring* genome sequence (RefSeq v1.0) made available by the International Wheat Genome Sequencing Consortium (IWGSC) in 2017 using default settings with the one exception that the number of times a GBS tag to be present and included for SNP calling was changed from the default value of 1 to 5 to increase the stringency in SNP calling. The identified SNPs with minor allele frequency (MAF) less than 5% and missing data more than 20% were removed from the analysis. All lines had missing data less than 20% and none of them were dropped due to missing percentage-filtering criterion.

## Population Structure and Genome-Wide Association Study Analysis

Population structure of 123 genotypes was assessed using Bayesian clustering algorithm in the program STRUCTURE v 2.3.4 [28] and principal component (PC) analysis using TASSEL [29] as described in Bhatta et al. [11].

Many GWAS studies were previously performed using mixed linear model (MLM), where population structure (Q) or PC was set as fixed effect and kinship (K) as a random effect to control false positives [29,30]. However, MLM may lead to confounding between population structure, kinship, and quantitative trait nucleotides (QTNs) that results in false negatives due to model overfitting [31]. Recently, the multi-locus mixed model (MLMM), which tests multiple markers simultaneously by fitting pseudo QTNs, in addition to testing markers in stepwise MLM, has been proposed, which is advantageous over conventional GLM and MLM testing one marker at a time [31]. One of the examples of recently popular GWAS analysis algorithm that is based on MLMM is FarmCPU (Fixed and random model Circulating Probability Unification) [31,32]. The FarmCPU uses a fixed effect model (FEM) and a random effect model (REM) iteratively to remove the confounding between testing markers and kinship that results in false negatives, prevents model over-fitting, and control false positives simultaneously [31]. Therefore, GWAS was performed on the adjusted BLUPs for each trait in each year to identify SNPs associated with grain yield and yield-related traits in SHWs using FarmCPU with population structure (Q<sub>1</sub> and Q<sub>2</sub>) or first three principal components (PC<sub>1</sub>, PC<sub>2</sub>, and PC<sub>3</sub>) as covariates by looking at the model fit using Quantile-Quantile (Q-Q) plots and FarmCPU calculated kinship [31] implemented in MVP R software package (<https://github.com/XiaoleiLiuBio/MVP>). A uniform suggestive genome wide significance threshold level of  $p$ -value =  $9.99E-05$  ( $-\log_{10}p=4.00$ ) was selected for MTAs considering the

deviation of the observed test statistics values from the expected test statistics values in the Q-Q plots [33,34] from the two-year results of this study.

### Haplotype Block Analysis

Haplotype blocks with linkage disequilibrium (LD) values (squared correlation coefficient between locus allele frequency;  $R^2 > 0.2$ ) in adjacent regions (< 500 Kb) of significant MTAs were visualized and plotted using default parameters (Hardy-Weinberg  $P$ -value cut off at 1% and MAF > 0.001) of Haploview software [35]. Phenotypic variance explained by each haplotype block on the trait of interest was calculated using multiple regression analysis that accounted for the population structure by removing the haplo-allele of less than 5% in SAS using PROC REG.

### Putative Candidate Gene Analysis

The genes underlying the MTAs and subsequently their annotations were retrieved using a Perl script and the IWGSC RefSeq v1.0 annotations provided for the Chinese Spring. The underlying genes were further examined for their association with grain yield and yield-related traits under drought stress using previously published literature. Additionally, the SnpEff program was used for SNP annotation and predicting the effects of SNPs on the protein function. The MTAs present within genes or are flanked (5K bases) by genes were investigated (<http://snpeff.sourceforge.net/>) [36].

## Results and Discussion

### Weather Conditions

The mean monthly air temperature was similar at Konya in both growing seasons with 13 °C in 2015-2016 and 12 °C in 2016-2017 compared to 25-year mean monthly air temperature (11 °C) in Turkey (Table 1). The total rainfall during 2016-2017 (243 mm) was slightly higher than 2015-2016 (222.4 mm) growing season. Total rainfall during the wheat-growing season (Sep-July) was 48.9% lower in 2015-16 and 44.2% lower in 2016-17 compared to 25-year mean total rainfall (435.1 mm) in Turkey. Although winter wheat water requirements are higher from mid-March to mid-June (from spring tillering period to mid-grain filling period), rainfall was very low in both years compared to 25-year mean rainfall. The plants were exposed to drought stress from tillering through grain-filling. Hence, results from this study can be used to understand the effects and genetics of drought in SHW.

### Phenotypic Variation for Yield and Yield-Related Traits

A combined analysis of variance (ANOVA) across years identified significant cross-over genotype x year interaction for all traits except for FLW and STMDIA (APPENDIX III).

Therefore, ANOVA was computed for both years separately and the results indicated that the SHWs showed significant variation for grain yield and yield-related traits in each year (Table 2). For instance, grain yield ranged from 200 to 341 g m<sup>-2</sup> with an average yield of 259 g m<sup>-2</sup> in 2016 and 241 to 392 g m<sup>-2</sup> with an average yield of 290 g m<sup>-2</sup> in 2017 (Table 2). The large variation among the traits in each year can be attributed to the collection of diverse accessions of SHWs from different countries and different genetic backgrounds [11,14].

Broad-sense heritability ( $H^2$ ) ranged from low to high heritability (Table 2). Low to moderate  $H^2$  was observed for GY (0.32—0.56), BMWT (0.39 —0.63), FLW (0.49 —0.67), and RTLN (0.31 —0.60), moderate  $H^2$  was observed for HI (0.63 —0.64) and STMDIA (0.57 —0.63), moderate to high  $H^2$  was observed for FLLN (0.53 —0.91) and FLA (0.52 —0.85), and high  $H^2$  was observed for TKW (0.75 —0.90) and GVWT (0.76 —0.91), indicating the genetic instability of these traits across years under drought stress. Similar  $H^2$  for most of these traits have been observed in previous studies [23,26,28,30,31,33,44–47].

#### Principal Component Analysis and Phenotypic Correlation

Principal component (PC) bi-plot analysis showed the association between grain yield and yield-related traits based on correlation matrix (Figure 1). The first two PCs that explained 43.4% (2016) to 44.9% (2017) of variation better explained the relationship between traits in two-dimensional space. In the PC-biplot, we observed two distinct groupings. The first one comprised of GY, HI, BMWT, TKW, GVWT, AWNLN, and RTLN whereas the second one had FLLN, FLW, and FLA (Figure 1). The traits grouping with GY are the more important traits for improving grain yield in drought-stressed conditions. The association observed in the PCA was supported by the significant correlation of grain yield with BMWT, HI, TKW, and GVW in both years (APPENDIX IV). Similar correlations for these traits were observed in the previous studies [21,22,34,37,40,42,43].

## Population Structure and Genome-Wide Association Study

Population structure analysis of 123 SHW was performed using 35,648 SNPs (MAF>0.05 and missing data <20%) using Bayesian clustering algorithm implemented in Structure software and the results showed that these lines were divided into three subgroups (APPENDIX V). The details of the population structure and genetic diversity of these SHWs has been previously reported in a detailed manner in Bhatta et al. [11].

The GWAS identified novel genomic regions for grain yield and yield-related traits and the MTA explained the high phenotypic variance. FarmCPU algorithm, with kinship, population structure (Q) or PC, BLUPs for each trait, and 35,648 GBS derived SNPs, was used to identify MTAs. The GBS derived SNPs were well distributed across each of the chromosome (APPENDIX VI). We identified 194 MTAs distributed across 21 chromosomes for grain yield and yield-related traits with phenotypic variance explained (PVE) ranging from 1.1 to 32.3% (Figures 2 and APPENDIX VII). The highest number of MTAs were observed for GY (29), followed by STMDIA (23), FLA (20), and TKW (20) while the lowest MTAs were observed for HI (10) (Figure 2). Of the 194 MTAs, 75 MTAs were detected on the A genome, 66 MTAs on the B genome, and 53 MTAs on the D genome. The highest the MTAs were present on chromosome 7A (26 MTAs) and the lowest MTAs on chromosome 3D (four MTAs) (Figure 2). Most of the MTAs identified in this study were year specific, suggesting the influence of genotype by environment interaction on the phenotype of the traits measured in the two years. However, 120 of the 194 significant SNPs were in 83 genes and 45 of these MTAs were present within genes, and their annotations suggested their potential role in drought stress. This result further provided confidence that the MTAs identified in the study are likely reliable MTAs (APPENDIX VII).

### *Grain Yield*

The 29 MTAs for GY were observed in 29 different genomic regions on seven chromosomes including 1B, 2B, 3A, 3D, 5B, 7A, and 7B with PVE ranging from 7.6 to 17.9% (**Figure 2** and APPENDIX VII). Earlier studies have reported QTLs/MTAs for grain yield on wheat chromosomes 1B [5,38,41,44], 2B [38–41,44], 3A [38,39,41,42], 3D [38], 5B [5,34,38,42,43,45], 7A [37,38,42,46], and 7B [38,41,42]. However, it is difficult to align our findings with previous studies due to the use of different marker systems [90K SNP, short sequence repeat (SSR), diversity arrays technology (DART) marker vs. GBS derived SNP marker), absence of precise location information in published studies, or the use of a different version of the reference wheat genome in previous studies than the latest IWGSC RefSeq v1.0. However, identification of several MTAs on same chromosome as earlier studies provided increased confidence on these associations.

The present study identified four major haplotype blocks (19 bp to 433 kbp) on chromosome 7A with two to six SNPs associated with grain yield in 2016 (**Figure 3**). First haplotype block consisted of six MTAs within 433 kbp range, second haplotype block consisted of four MTAs within 81 bp range, third haplotype block consisted of two MTAs within 19 bp range, and fourth haplotype block consisted of three MTAs within 314 kbp range. The PVE on grain yield by the first, second, third, and fourth haplotype blocks were 17.2%, 24.6%, 21.9%, and 8.2%, respectively.

One MTA (S7A\_112977027; 112977027 bp) present in between second (537 kb away) and third (837 kb) haplotype block was within the gene, TraesCS7A01G158200.1, and PVE on grain yield was 12.8% (APPENDIX VII). This gene was annotated as a member of sentrin-



specific protease of *Ulp1* (Ubiquitin-like Protease 1) gene family (**Table 3**). The Ulp1 is a small ubiquitin related modifier (SUMO) specific protease that affects several important biological processes in plants including response to abiotic stress [47]. It has been shown to have a role in drought tolerance in *Arabidopsis* (*Arabidopsis thaliana*) [48] and rice (*Oryza sativa*) [49,50]. This makes this MTA interesting and a stronger candidate for future functional validation studies.

Another major haplotype block (18 kbp) of three MTAs was observed on chromosome 3A in 2017 (Figure 3) with PVE on grain yield by this haplotype block was 13.1% (**Figure 3**). The chromosome 3A is known to be an important chromosome that contains useful QTLs for grain yield and yield-related traits [38,39,41,42,51] and the haplotype block identified will have a significance in the crop improvement program. All three MTAs present in this haplotype block of chromosome 3A were found in the gene, TraesCS3A01G047300 (APPENDIX VII), which was annotated as a member of the F-box gene family (**Table 4**). These three SNPs were indicated as having a moderate impact on the protein as they resulted into a missense mutation and caused an amino acid change. Such changes may alter the function of of the protein [36], which makes this F-box gene a strong candidate for future functional characterization studies under drought tolerance in wheat. The F-box proteins is known to regulate many important biological processes such as embryogenesis, floral development, plant growth and development, biotic and abiotic stress, hormonal responses, and senescence [52]. Two other MTAs observed on chromosome 3A and 3D were present within genes (F-box family protein: TraesCS3A01G445100 and disease resistance protein RPM1: TraesCS3D01G002700) that had been previously reported be involved in drought tolerance [52,53] (**Table 4**).

The grain yield haplotype blocks and other MTAs identified in this study have not been mapped to date and four MTAs were in the genes whose functional annotations suggest that they are likely involved in drought tolerance. This result implied that haplotype blocks observed on chromosome 3A (3 MTAs) and 7A (16 MTAs), and one MTA on chromosome 3D (1) for grain yield are novel and may potentially be used in a marker-assisted breeding program focusing on improving drought tolerance in wheat after validating them in different populations and environments.

#### *Harvest Index*

A total of 10 SNPs significantly associated with HI were identified on chromosomes 1D, 2A, 2D, 3A, 3D, 5B, 6B, 6D, and 7B (**Figure 2**) with PVE ranging from 2.2 to 18.7% (APPENDIX VII). Previous studies have reported QTLs/MTAs responsible for HI on chromosomes 2D [37], 3A [37,42], 6B [54], and 7B [42]. The six MTAs identified for HI on chromosomes 1D, 2A, 3D, 5B, and 6D have not been reported, to the best of our knowledge, and they are potentially novel MTAs responsible for HI.

Six MTAs for HI detected on chromosomes 2A, 3A, 6B, 6D, and 7B were found within genes and two of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and S5**). The two genes are WRKY transcription factor (TraesCS3A01G343700) found on chromosome 3A and cytochrome P450 (TraesCS6D01G170900.1) found on chromosome 6D. The role of WRKY transcription factor is well known in abiotic stress including drought tolerance [55,56]. The cytochrome P450 genes are a large superfamily of enzymes and are involved in many metabolic pathways including drought tolerance in rice [57]

and Arabidopsis [58,59]. The multi-trait marker associated with GY and HI was located on chromosome 5B (S5B\_598463062) with PVE ranging from 15.9% to 18.7% (APPENDIX VII)

#### *Biomass Weight*

The 15 MTAs responsible for BMWT were identified on chromosomes 1D, 2B, 3A, 4A, 6D, and 7B (**Figure 2**) with PVE ranging from 4.9 to 14.4% (APPENDIX VII). Previous studies had reported QTLs/MTAs responsible for BMWT on chromosomes 1D [42,54], 2B [37], 6D [37], and 7B [42]. The four MTAs identified for BMWT on chromosome 3A and 4A have not been reported and they are potentially novel MTAs responsible for BMWT.

A novel haplotype block (38 kbp) of three SNPs on chromosome 3A associated with BMWT was identified in 2017 (**Figure 3**) with PVE by the haplotype block was 11.7%. This MTA (S3A\_25012018) was also associated with grain yield and PVE ranged from 12.7% (GY) to 14.4% (BMWT).

All three MTAs present in this haplotype block were within genes (APPENDIX VII) and one of the genes had annotations suggesting its involvement in drought tolerance was an F-box family protein (TraesCS3A01G047300) (**Table 4**) [52]. Excluding MTAs on haplotype block, eight MTAs for BMWT detected on chromosomes 1D, 2B, 6D, and 7B were found within genes (APPENDIX VII) and two of the genes had annotations suggesting its involvement in drought stress (**Tables 3 and 4**). The genes associated with two MTAs are F-box family protein (TraesCS7B01G242600) [52] and protein DETOXIFICATION containing multi-antimicrobial extrusion protein (MatE) (TraesCS1D01G357500) [60].

#### *Thousand Kernel Weight*

A total of 20 MTAs responsible for TKW were detected in 19 different genomic regions on chromosomes 1A, 2A, 2B, 2D, 3A, 3B, 4A, 4B, 4D, 5B, 6D, 7B, and 7D (**Figure 2**) with PVE ranging from 1.6 to 22.2% (APPENDIX VII). Earlier studies have reported QTLs/MTAs for TKW on chromosomes 1A [40,42,44,61], 2A [46], 2B [40,42,46], 2D [44], 3A [40,61,62], 3B [46,63], 4B [5], 5B [61], 7B [42], and 7D [46]. In the present study, only one MTA (S2D\_7309581) for TKW was detected on chromosome 2D in both years and assumed to be a stable MTA because this MTA was detected despite significant genotype x year interaction. The five MTAs identified for TKW on chromosomes 4A, 4D, and 6D have not been previously reported and they are potentially novel MTAs responsible for TKW.

Twelve MTAs for TKW detected on chromosomes 2A, 2B, 3A, 3B, 4A, 4B, 5B, and 6D were found within genes (APPENDIX VII) and five of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with MTAs involved in drought tolerance are F-box family protein (chromosome 3A; TraesCS3A01G047300) [52], protein kinase family protein (chromosome 4A; TraesCS4A01G347600 and chromosome 6D; TraesCS6D01G360800) [64], cytochrome P450 family protein (chromosome 4D; TraesCS4D01G364700) [57–59], and zinc finger (C3HC4-type RING finger) family protein (chromosome 4B; TraesCS4B01G344200.1) [65–67]. The SNPs S4D\_509427923 was annotated as a missense variant and thus may have a moderate impact on the protein function (APPENDIX VII).

#### *Grain Volume Weight*

The 13 MTAs responsible for GVWT were identified on chromosomes 1A, 2A, 2B, 2D, 3A, 4A, 5A, 6A, and 7A (**Figure 2**) with PVE ranging from 1.3 to 16.2% (APPENDIX VII).

Earlier studies have reported QTLs/MTAs for GVWT on chromosomes 1A [38], 2A [38,68], 2B [38,68], 2D [38,68], 5A [38] and 7A [38,68]. The four MTAs identified for GVWT on chromosomes 3A, 4A, and 6A have not been previously reported and they are potentially novel MTAs responsible for GVWT.

Eight MTAs responsible for GVWT detected on chromosomes 1A, 2A, 2B, 4A, 6A, and 7A were found in the gene regions (APPENDIX VII) and two of these genes have annotations suggesting their involvement in drought stress (**Table 4**). The genes associated with two MTAs involved in drought tolerance are cytochrome P450 (TraesCS1A01G334800) [57–59] on chromosome 1A and microtubule associated protein family protein (TraesCS4A01G074200.2) [69] on chromosome 4A.

#### *Awn Length*

The 20 MTAs responsible for AWNLN were observed on chromosomes 1D, 2A, 2B, 3B, 4A, 4B, 4D, 5A, 5B, 5D, 6B, and 7A (**Figure 2**) with PVE ranging from 1.1 to 20.1% (APPENDIX VII). Earlier studies have reported QTLs/MTAs for AWNLN on chromosomes 2A [70,71], 4A [71], 4B [71], 5A [71], and 6B [70,71]. The nine MTAs identified for AWNLN on chromosomes 2B, 3B, 4D, 5B, 5D, and 7A have not been previously reported and they are potentially novel MTAs responsible for AWNLN.

Eleven MTAs for AWNLN detected on chromosomes 1D, 2A, 4D, 5A, 5B, 6B, and 7A were found in the gene regions (APPENDIX VII) and four of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with four MTAs involved in drought tolerance are 60S ribosomal protein L18a (chromosome 4D; TraesCS4D01G290700) [72], guanine nucleotide exchange family protein (chromosome 5A;

TraesCS5A01G361300) [73], and F-box family protein (chromosome 5B; TraesCS5B01G038700 and chromosome 6B; TraesCS6B01G001000) [52]. It has been reported that the putative 60S ribosomal protein L18a is an up-regulated transcripts in response to drought stress in ears and silks during the flowering stage in maize [72].

#### *Flag Leaf Length*

The 13 MTAs responsible for FLLN were detected on chromosomes 1B, 1D, 2A, 2B, 2D, 4A, 6D, and 7B (**Figure 2**) with PVE ranging from 1.58 to 32.3% (APPENDIX VII). Previous studies have reported QTLs for FLLN on chromosomes 1B [74,75], 2B [19,74–76], 2D [20,74], 4A [19,74,75], and 7B [20]. The four MTAs identified for FLLN on chromosomes 1D, 2A, and 6D have not been previously reported and they are potentially novel MTAs responsible for FLLN.

Eleven MTAs responsible for FLLN detected on chromosomes 1B, 1D, 2B, 2D, 4A, 6D, and 7B were found in the gene regions (APPENDIX VII) and four of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with four MTAs involved in drought stress are F-box family protein (chromosome 4A: TraesCS4A01G325200) [52], cytochrome P450 (chromosome 2B; TraesCS2B01G167500 and chromosome 6D; TraesCS6D01G386300) [57–59], and Rp1-like protein (chromosome 1B; TraesCS1B01G400600) [53].

#### *Flag Leaf Width*

The 16 MTAs responsible for FLW were detected on chromosomes 1A, 1B, 1D, 2B, 2D, 4B, 6B, and 6D (**Figure 2**) with PVE ranging from 1.6 to 15.2% (APPENDIX VII). Previous

studies have found QTLs for FLW on chromosomes 1B [20,74,76], 1D [19,74], 2B [19,74], 2D [19,20,74], 4B [19,76], and 6B [19,74,75]. The two MTAs identified for FLW on chromosomes 1A and 6D have not been previously reported and they are potentially novel MTAs responsible for FLW.

Thirteen MTAs for FLW detected on chromosomes 1A, 1B, 1D, 2D, 4B, 6B, and 6D were found in the gene regions (APPENDIX VII) and three of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with three MTAs involved in drought stress are citrate-binding protein (chromosome 1A; TraesCS1A01G326700) [77], F-box family protein (chromosome 6B; TraesCS6B01G042800) [52], and mitochondrial transcription termination factor-like (chromosome 6D; TraesCS6D01G040100) [78]. The SNPs S1A\_516732460 and S6D\_16376439 were annotated as a missense variant and thus may impact the function of the proteins that are annotated as citrate-binding protein and mitochondrial transcription termination factor-like protein, respectively (APPENDIX VII).

#### *Flag Leaf Area*

The 20 MTAs for FLW were detected on chromosomes 1A, 1B, 1D, 2A, 2D, 4D, 5A, 6B, and 7D (**Figure S2**) with PVE ranging from 8.1 to 23.1% (APPENDIX VII). Previous studies have reported QTLs for FLA on chromosomes 1B [19,75], 1D [19,20,74], 2A [19,20,74], 2D [19,20,74], 4D [75], 5A [20,74–76], 6B [75], and 7D [20]. The four MTAs identified for FLA on chromosome 1A have not been previously reported and they are potentially novel MTAs responsible for FLA.

Three novel haplotype blocks were observed for FLA on chromosomes 1A (two MTAs), 6B (two MTAs) and 7D (three MTAs) with PVE by these haplotype block ranging from 5.5% to 8.6% (**Figure 3**). Fourteen MTAs for FLA detected on chromosomes 1A, 1B, 1D, 2A, 4D, 5A, 6B, and 7D were found in the gene regions (APPENDIX VII) and five of these genes had annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with five MTAs involved in drought stress are citrate-binding protein (TraesCS1A01G326700), P-loop containing nucleoside triphosphate hydrolases superfamily protein (TraesCS1D01G197200) [79], cytochrome P450 (TraesCS6B01G125900) [57–59], and NBS-LRR resistance like protein [53].

The multi-trait marker associated with FLW and FLA was located on chromosome 1A (S1A\_516732460) with PVE ranging from 8.0% to 9.5% (**Table S4**). Another multi-trait marker associated with FLLN and FLA was located on chromosome 2A (S2A\_29874199) with PVE ranging from 23.1 to 32.3% (APPENDIX VI). The multi-trait MTA indicates that the related candidate gene may affect multiple traits.

#### *Stem Diameter*

In the present study, 23 MTAs responsible for STMDIA were identified on chromosomes 1A, 1D, 2B, 2D, 3A, 3B, 3D, 4D, 5A, 5B, 6A, 6B, 6D, 7A, and 7B (**Figure 2**) with PVE ranging from 2.7 to 28.8% (APPENDIX VII). Earlier study has identified one minor QTL (Qsd-3B) for STMDIA on chromosome 3B that explained 8.7% of the phenotypic variance [80]. That means, the 19 MTAs detected on chromosomes 1A, 1D, 2B, 2D, 3A, 3D, 4D, 5A, 5B, 6A, 6B, 6D, 7A, and 7B in the present study except chromosome 3B may potentially be a novel MTAs controlling STMDIA under drought stress.



Four MTAs were detected on chromosome 3B and two of them (1 bp apart) were observed in one haplotype block in 2016 with PVE was 9.2% (**Figure 3**). Fifteen MTAs for STMDIA detected on chromosomes 1D, 2B, 3A, 3B, 6B, 6D, 7A, and 7B were found in the gene regions (APPENDIX VII) and four of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with four MTAs involved in drought stress are leucine-rich repeat receptor-like protein kinase family protein (TraesCS3D01G028500) [81], protein kinase (TraesCS6A01G122200.1) [64], disease resistance protein (NBS-LRR class) family (TraesCS1D01G341500 and TraesCS6B01G346900 ) [53], and F-box protein family (TraesCS6B01G347000) [52].

#### *Root Length*

Root length is one of the most important traits under drought stress. We have measured root length 3-4 days after anthesis (Zadoks 60 growth stage) under drought stressed field condition using WinRhizo® (WinRhizo reg. 2009c, Regent Instruments Inc., Quebec City, Quebec, Canada). This trait is very unique compared to previous studies where they focused on the roots of seedlings [82–85] rather than direct field-based measurements (labor intensive, time consuming, and expensive). Identification of QTL governing root length is very important in wheat especially for the wheat grown under drought stress. Limited information is available on QTL related to root traits in wheat [82,84–87].

In the present study, 15 MTAs responsible for RTLN were identified on chromosomes 2B, 2D, 3B, 5B, 6A, 6D, and 7A (**Figure 2**) with PVE ranging from 5.3% to 18.5% (APPENDIX VII). Earlier studies have reported QTLs for RTLN on chromosomes 2B [83], 3B [84], 5B [82,84], 6A [83,84], and 6D [82,85] in hexaploid wheat and on chromosomes 2B [86,87], 3B

[87], 6A [87], and 7A [87] in tetraploid wheat. The MTA identified for RTLN on chromosome 2D has not been previously reported and it is potentially novel MTAs responsible for RTLN under drought stress. Furthermore, previous studies identified very few QTLs for RTLN on the D genome of wheat [82,85]. Therefore, the MTAs (eight MTAs) for RTLN detected on the D genome of SHWs in this study are potentially novel.

Seven out of eight MTAs responsible for RTLN were present on chromosome 6D. Two haplotype blocks (haplotype block1 of size 64 kbp and haplotype block2 of size 5kbp) were identified from five out of seven MTAs for RTLN on chromosome 6D with PVE ranging from 5.0% to 11.8% (**Figure 3**). One SNP (S6D\_435300571) present in the haplotype block2 was found in the gene (TraesCS6D01G332800) with PVE was 13.0%. The gene associated with this SNP is protein detoxification gene containing multi-antimicrobial extrusion protein (MatE) (**Table 3**) and has been reported to be expressed mainly in the root than shoots under drought stress [88]. For instance, MatE family genes such as *HvAACT1* in barley [89] and *TaMate* in wheat [90], encode proteins that are primarily localized to root epidermis cells [89] and required for external resistance [60]. In the present study, this gene was also significantly associated with BMWT on chromosome 1D. This result implies that that this gene plays an important role for RTLN and BMWT in drought-stressed conditions.

Excluding MTAs on haplotype block, eight MTAs for RTLN detected on chromosomes 2D, 3B, 5B, 6A, and 7A were found in the gene regions (APPENDIX VII) and four of these genes have annotations suggesting their involvement in drought stress (**Tables 3 and 4**). The genes associated with four MTAs involved in drought stress are GRAM domain-containing protein/ABA-responsive (TraesCS5B01G502200) [91,92][91,92][92,93][90,91], phosphatase 2C family protein (TraesCS7A01G143200.2)[93], and disease resistance protein RPM1

(TraesCS2D01G541000.1) [53]. The SNPs S7A\_94404310 was annotated as a missense variant and thus may have a moderate impact on the protein function (APPENDIX VII).

#### Potential Candidate Gene Annotations Affecting Yield and Yield-Related Traits Under Drought Stress

This study identified ~194 MTAs present on different chromosomes and associated with multiple traits. Of these 62 MTAs were either associated with same trait in multiple years (MTA stability in different environments) or multiple traits within the same year or across years (suggesting epistasis) (Table S5). Additionally, ~45 of the MTAs were present in genes with annotations relevant to the respective trait under drought stress (Tables 3 and 4). Interestingly, we noticed MTAs associated with same or related traits were located within genes that had the exact same annotation (**Figure 4**). For instance, some of the MTAs for GY (2 MTAs), BMWT (2), TKW (1), AWNLN (2), FLLN (1), FLW (1), and STMDIA (1) were located within genes annotated as F-box family protein. Similarly, the genes annotated as cytochrome P450 harbored MTAs for HI (1), TKW (1), FLA (2), GVWT (1), and FLLN (2). Additional examples are provided in **Figure 4**. This result indicates the likely gene families that are important for grain yield and grain-yield related traits under drought stress.

#### Conclusions

The present study showed SHWs have large amounts of genetic variation for grain yield and yield-related traits. The GWAS in 123 SHWs using 35,648 SNPs identified several novel (90 MTAs: 45 MTAs on the A genome, 11 on the B genome, and 34 on the D-genome) genomic regions or haplotype blocks associated with grain yield and yield-related traits in drought

stressed condition. Most of the MTAs (120 MTAs) were present in genes and several of them (45 MTAs) were annotated with functions related to drought stress. This provided further evidence for the reliability of the MTAs identified. We also identified MTAs on different chromosomes associated with multiple traits but within genes having the same annotation. This resulted in the identification of candidate genes belonging to the same gene family that likely have a major role in affecting yield and yield-related traits under drought stress in SHWs. The large number of MTAs especially on the D genome (53 MTAs with 34 MTAs being novel) identified in this study demonstrate the potential of SHWs for elucidating the genetic architecture of complex traits and provide an opportunity for further improvement of wheat under rapidly growing drought stressed environment throughout the world.

#### Abbreviations

ANOVA	Analysis of Variance
AWNLN	Awn Length
BLUP	Best Linear Unbiased Prediction
BMWT	Biomass Weight
FarmCPU	Fixed and random model Circulating Probability Unification
FLA	Flag Leaf Area
FLLN	Flag Leaf Length
FLW	Flag Leaf Width
GBS	Genotyping-By-Sequencing
GVWT	Grain Volume Weight
GWAS	Genome Wide Association Study
GY	Grain Yield
HI	Harvest Index
IWGC	International Wheat Genome Sequencing Consortium

MAF	Minor Allele Frequency
QTN	Quantitative Trait Nucleotide
MLM	Mixed Linear Model
MLMM	Multi-locus Mixed Model
MTA	Marker Trait Association
QTL	Quantitative Trait Loci
RTLN	Root Length
SHW	Synthetic Hexaploid Wheat
SNP	Single Nucleotide Polymorphism
STMDIA	Stem Diameter
TKW	Thousand Kernel Weight

## References

1. Kang, Y.; Khan, S.; Ma, X. Climate change impacts on crop yield, crop water productivity and food security - A review. *Prog. Nat. Sci.* **2009**, *19*, 1665–1674, doi:10.1016/j.pnsc.2009.08.001.
2. Becker, S. R.; Byrne, P. F.; Reid, S. D.; Bauerle, W. L.; McKay, J. K.; Haley, S. D. Root traits contributing to drought tolerance of synthetic hexaploid wheat in a greenhouse study. *Euphytica* **2016**, *207*, 213–224, doi:10.1007/s10681-015-1574-1.
3. Gupta, P.; Balyan, H.; Gahlaut, V.; Gupta, P. K.; Balyan, H. S.; Gahlaut, V. QTL analysis for drought tolerance in wheat: Present status and future possibilities. *Agronomy* **2017**, *7*, 5, doi:10.3390/agronomy7010005.
4. Smith, A. B.; Matthews, J. L. Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. *Nat. Hazards* **2015**, *77*, 1829–1851, doi:10.1007/s11069-015-1678-x.
5. Pinto, R. S.; Reynolds, M. P.; Mathews, K. L.; McIntyre, C. L.; Olivares-Villegas, J.-J.; Chapman, S. C. Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theor. Appl. Genet.* **2010**, *121*, 1001–1021, doi:10.1007/s00122-010-1351-4.
6. Gummadov, N.; Keser, M.; Akin, B.; Cakmak, M.; Mert, Z.; Taner, S.; Ozturk, I.; Topal, A.; Yazar, S.; Morgounov, A. Genetic gains in wheat in Turkey: Winter wheat for irrigated conditions. *Crop J.* **2015**, *3*, 507–516, doi:10.1016/J.CJ.2015.07.007.

7. Cavanagh, C. R.; Chao, S.; Wang, S.; Huang, B. E.; Stephen, S.; Kiani, S. et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci.* **2013**, *110*, 8057–8062, doi:10.1073/pnas.1217133110.
8. Cox, T. Deepening the wheat gene pool. *J. Crop Prod.* **1997**, *1*, 145–168, doi:10.1300/J144v01n01.
9. Dreisigacker, S.; Kishii, M.; Lage, J.; Warburton, M. Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. In *Australian Journal of Agricultural Research*; 2008; Vol. 59, pp. 413–420.
10. Ogonnaya, F. C.; Abdalla, O.; Mujeeb-Kazi, A.; Kazi, A. G.; Xu, S. S.; Gosman, N.; Lagudah, E. S.; Bonnett, D.; Sorrells, M. E.; Tsujimoto, H. Synthetic hexaploids: Harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.* **2013**, *37*, 35–122, doi:10.1002/9781118497869.ch2.
11. Bhatta, M.; Morgounov, A.; Belamkar, V.; Poland, J.; Baenziger, P. S. Unlocking the novel genetic diversity and population structure of synthetic hexaploid wheat. *BMC Genomics* **2018**, *19*, 591, doi:10.1186/s12864-018-4969-2.
12. Zegeye, H.; Rasheed, A.; Makdis, F.; Badebo, A.; Ogonnaya, F. C. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One* **2014**, *9*, e105593, doi:10.1371/journal.pone.0105593.
13. Das, M. K.; Bai, G.; Mujeeb-Kazi, A.; Rajaram, S. Genetic diversity among synthetic hexaploid wheat accessions (*Triticum aestivum*) with resistance to several fungal diseases. *Genet. Resour. Crop Evol.* **2016**, *63*, 1285–1296, doi:10.1007/s10722-015-0312-9.

14. Morgounov, A.; Abugalieva, A.; Akan, K.; Akın, B.; Baenziger, S.; Bhatta, M.; Dababat, A. A.; Demir, L.; Dutbayev, Y.; El Bouhssini, M.; Erginbaş-Orakci, G.; Kishii, M.; Keser, M.; Koç, E.; Kurespek, A.; Mujeeb-Kazi, A.; Yorgancılar, A.; Özdemir, F.; Öztürk, I.; Payne, T.; Qadimaliyeva, G.; Shamanin, V.; Subasi, K.; Suleymanova, G.; Yakişir, E.; Zelenskiy, Y. High-yielding winter synthetic hexaploid wheats resistant to multiple diseases and pests. *Plant Genet. Resour.* **2018**, 16(3), 273-278, doi:10.1017/S147926211700017X.
15. Becker, S. R.; Byrne, P. F.; Reid, S. D.; Bauerle, W. L.; McKay, J. K.; Haley, S. D. Root traits contributing to drought tolerance of synthetic hexaploid wheat in a greenhouse study. *Euphytica* **2016**, 207, 213–224, doi:10.1007/s10681-015-1574-1.
16. Ogonnaya, F. C.; Imtiaz, M.; Bariana, H. S.; McLean, M.; Shankar, M. M.; Hollaway, G. J.; Trethowan, R. M.; Lagudah, E. S.; Van Ginkel, M. Mining synthetic hexaploids for multiple disease resistance to improve bread wheat. In *Australian Journal of Agricultural Research*; 2008; Vol. 59, pp. 421–431.
17. Jighly, A.; Alagu, M.; Makdis, F.; Singh, M.; Singh, S.; Emebiri, L. C.; Ogonnaya, F. C. Genomic regions conferring resistance to multiple fungal pathogens in synthetic hexaploid wheat. *Mol. Breed.* **2016**, 36, 1–19, doi:10.1007/s11032-016-0541-4.
18. Evaluation of deficit irrigation for efficient sheep production from permanent sown pastures in a dry continental climate. *Agric. Water Manag.* **2013**, 119, 135–143, doi:10.1016/j.agwat.2012.12.017.
19. Fan, X.; Cui, F.; Zhao, C.; Zhang, W.; Yang, L.; Zhao, X.; Han, J.; Su, Q.; Ji, J.; Zhao, Z.; Tong, Y.; Li, J. QTLs for flag leaf size and their influence on yield-related traits in wheat (*Triticum aestivum* L.). *Mol. Breed.* **2015**, 35, 24, doi:10.1007/s11032-015-0205-9.



20. Hussain, W.; Baenziger, P. S.; Belamkar, V.; Guttieri, M. J.; Venegas, J. P.; Easterly, A.; Sallam, A.; Poland, J. Genotyping-by-sequencing derived high-density linkage map and its application to qtl mapping of flag leaf traits in bread wheat. *Sci. Rep.* **2017**, *7*, 16394, doi:10.1038/s41598-017-16006-z.
21. Bhatta, M.; Eskridge, K. M.; Rose, D. J.; Santra, D. K.; Baenziger, P. S.; Regassa, T. Seeding rate, genotype, and topdressed nitrogen effects on yield and agronomic characteristics of winter wheat. *Crop Sci.* **2017**, *57*, 951–963, doi:10.2135/cropsci2016.02.0103.
22. Bhatta, M.; Regassa, T.; Rose, D. J.; Baenziger, P. S.; Eskridge, K. M.; Santra, D. K.; Poudel, R. Genotype, environment, seeding rate, and top-dressed nitrogen effects on end-use quality of modern Nebraska winter wheat. *J. Sci. Food Agric.* **2017**, *97(15)*, 5311–5318.
23. SAS 9.4 Product Documentation Available online: <https://support.sas.com/documentation/94/> (accessed on Aug 16, 2018).
24. Kassambara, A.; Mundt, F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses; 2017. Available online: <http://www.sthda.com/english/rpkgs/factoextra/> (accessed on Aug 16, 2018).
25. Poland, J. A.; Brown, P. J.; Sorrells, M. E.; Jannink, J. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0032253.
26. Glaubitz, J. C.; Casstevens, T. M.; Lu, F.; Harriman, J.; Elshire, R. J.; Sun, Q.; Buckler, E. S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **2014**, *9*, doi:10.1371/journal.pone.0090346.

27. Pritchard, J. K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
28. Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y.; Buckler, E. S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635, doi:10.1093/bioinformatics/btm308.
29. Yu, J.; Pressoir, G.; Briggs, W. H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J. F.; McMullen, M. D.; Gaut, B. S.; Nielsen, D. M.; Holland, J. B.; Kresovich, S.; Buckler, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208, doi:10.1038/ng1702.
30. Zhang, Z.; Ersoz, E.; Lai, C.-Q.; Todhunter, R. J.; Tiwari, H. K.; Gore, M. A.; Bradbury, P. J.; Yu, J.; Arnett, D. K.; Ordovas, J. M.; Buckler, E. S. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360, doi:10.1038/ng.546.
31. Liu, X.; Huang, M.; Fan, B.; Buckler, E. S.; Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genet.* **2016**, *12*, e1005767, doi:10.1371/journal.pgen.1005767.
32. Arora, S.; Singh, N.; Kaur, S.; Bains, N. S.; Uauy, C.; Poland, J.; Chhuneja, P. Genome-wide association study of grain architecture in wild wheat *aegilops tauschii*. *Front. Plant Sci.* **2017**, *8*, doi:10.3389/fpls.2017.00886.
33. Sukumaran, S.; Xiang, W.; Bean, S. R.; Pedersen, J. F.; Kresovich, S.; Tuinstra, M. R.; Tesso, T. T.; Hamblin, M. T.; Yu, J. Association mapping for grain quality in a diverse sorghum collection. *Plant Genome J.* **2012**, *5*, 126, doi:10.3835/plantgenome2012.07.0016.

34. Sukumaran, S.; Dreisigacker, S.; Lopes, M.; Chavez, P.; Reynolds, M. P. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* **2015**, *128*, 353–363, doi:10.1007/s00122-014-2435-3.
35. Barrett, J. C.; Fry, B.; Maller, J.; Daly, M. J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265, doi:10.1093/bioinformatics/bth457.
36. Cingolani, P.; Platts, A.; Wang, L. L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S. J.; Lu, X.; Ruden, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **2012**, *6*, 80–92, doi:10.4161/fly.19695.
37. Kumar, N.; Kulwal, P. L.; Balyan, H. S.; Gupta, P. K. QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. *Mol. Breed.* **2007**, *19*, 163–177, doi:10.1007/s11032-006-9056-8.
38. Bordes, J.; Goudemand, E.; Duchalais, L.; Chevarin, L.; Oury, F. X.; Heumez, E.; Lapiere, A.; Perretant, M. R.; Rolland, B.; Beghin, D.; Laurent, V.; Le Gouis, J.; Storlie, E.; Robert, O.; Charmet, G. Genome-wide association mapping of three important traits using bread wheat elite breeding populations. *Mol. Breed.* **2014**, *33*, 755–768, doi:10.1007/s11032-013-0004-0.
39. Hoffstetter, A.; Cabrera, A.; Sneller, C. Identifying Quantitative Trait Loci for Economic Traits in an Elite Soft Red Winter Wheat Population. *Crop Sci.* **2016**, *56*, 547, doi:10.2135/cropsci2015.06.0332.
40. Lozada, D. N.; Mason, R. E.; Babar, M. A.; Carver, B. F.; Guedira, G. B.; Merrill, K.; Arguello, M. N.; Acuna, A.; Vieira, L.; Holder, A.; Addison, C.; Moon, D. E.; Miller, R. G.;

Dreisigacker, S. Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. *Euphytica* **2017**, *213*, doi:10.1007/s10681-017-2005-2.

41. Sehgal, D.; Autrique, E.; Singh, R.; Ellis, M.; Singh, S.; Dreisigacker, S. Identification of genomic regions for grain yield and yield stability and their epistatic interactions. *Sci. Rep.* **2017**, *7*, doi:10.1038/srep41578.

42. Neumann, K.; Kobiljski, B.; Denčić, S.; Varshney, R. K.; Börner, A. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* **2011**, *27*, 37–58, doi:10.1007/s11032-010-9411-7.

43. Edae, E. A.; Byrne, P. F.; Haley, S. D.; Lopes, M. S.; Reynolds, M. P. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* **2014**, *127*, 791–807, doi:10.1007/s00122-013-2257-8.

44. Ogbonnaya, F. C.; Rasheed, A.; Okechukwu, E. C.; Jighly, A.; Makdis, F.; Wuletaw, T.; Hagra, A.; Uguru, M. I.; Agbo, C. U. Genome-wide association study for agronomic and physiological traits in spring wheat evaluated in a range of heat prone environments. *Theor. Appl. Genet.* **2017**, *130*, 1819–1835, doi:10.1007/s00122-017-2927-z.

45. Wang, S.-X.; Zhu, Y.-L.; Zhang, D.-X.; Shao, H.; Liu, P.; Hu, J.-B.; Zhang, H.; Zhang, H.-P.; Chang, C.; Lu, J.; Xia, X.-C.; Sun, G.-L.; Ma, C.-X. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLOS ONE* **2017**, *12*, e0188662, doi:10.1371/journal.pone.0188662.

46. Sukumaran, S.; Lopes, M.; Dreisigacker, S.; Reynolds, M. Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for

grain weight and grain number. *Theor. Appl. Genet.* **2018**, *131*, 985–998, doi:10.1007/s00122-017-3037-7.

47. Murtas, G. A Nuclear Protease Required for Flowering-Time Regulation in Arabidopsis Reduces the Abundance of SMALL UBIQUITIN-RELATED MODIFIER Conjugates. *Plant Cell* **2003**, *15*, 2308–2319, doi:10.1105/tpc.015487.

48. Catala, R.; Ouyang, J.; Abreu, I. A.; Hu, Y.; Seo, H.; Zhang, X.; Chua, N.-H. The Arabidopsis E3 SUMO Ligase SIZ1 Regulates Plant Growth and Drought Responses. *Plant Cell* **2007**, *19*, 2952–2966, doi:10.1105/tpc.106.049981.

49. Choudhary, M. K.; Basu, D.; Datta, A.; Chakraborty, N.; Chakraborty, S. Dehydration-responsive Nuclear Proteome of Rice (*Oryza sativa* L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. *Mol. Cell. Proteomics* **2009**, *8*, 1579–1598, doi:10.1074/mcp.M800601-MCP200.

50. Park, H. J.; Kim, W.-Y.; Park, H. C.; Lee, S. Y.; Bohnert, H. J.; Yun, D.-J. SUMO and SUMOylation in plants. *Mol. Cells* **2011**, *32*, 305–316, doi:10.1007/s10059-011-0122-7.

51. Ali, M. L.; Baenziger, P. S.; Ajlouni, Z. A.; Campbell, B. T.; Gill, K. S.; Eskridge, K. M.; Mujeeb-Kazi, A.; Dweikat, I. Mapping QTL for agronomic traits on wheat chromosome 3A and a comparison of recombinant inbred chromosome line populations. *Crop Sci.* **2011**, *51*, 553–566, doi:10.2135/cropsci2010.06.0359.

52. Lechner, E.; Achard, P.; Vansiri, A.; Potuschak, T.; Genschik, P. F-box proteins everywhere. *Curr. Opin. Plant Biol.* **2006**, *9*, 631–638, doi:10.1016/j.pbi.2006.09.003.

53. Lee, H.-A.; Yeom, S.-I. Plant NB-LRR proteins: tightly regulated sensors in a complex manner. *Brief. Funct. Genomics* **2015**, *14*, 233–242, doi:10.1093/bfgp/elv012.
54. Ain, Q.; Rasheed, A.; Anwar, A.; Mahmood, T.; Imtiaz, M.; Mahmood, T.; Xia, X.; He, Z.; Quraishi, U. M. Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. *Front. Plant Sci.* **2015**, *6*, doi:10.3389/fpls.2015.00743.
55. He, G.-H.; Xu, J.-Y.; Wang, Y.-X.; Liu, J.-M.; Li, P.-S.; Chen, M.; Ma, Y.-Z.; Xu, Z.-S. Drought-responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat confer drought and/or heat resistance in Arabidopsis. *BMC Plant Biol.* **2016**, *16*, 116, doi:10.1186/s12870-016-0806-4.
56. Ning, P.; Liu, C.; Kang, J.; Lv, J. Genome-wide analysis of WRKY transcription factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit condition. *PeerJ* **2017**, *5*, e3232, doi:10.7717/peerj.3232.
57. Tamiru, M.; Undan, J. R.; Takagi, H.; Abe, A.; Yoshida, K.; Undan, J. Q.; Natsume, S.; Uemura, A.; Saitoh, H.; Matsumura, H.; Urasaki, N.; Yokota, T.; Terauchi, R. A cytochrome P450, OsDSS1, is involved in growth and drought stress responses in rice (*Oryza sativa* L.). *Plant Mol. Biol.* **2015**, *88*, 85–99, doi:10.1007/s11103-015-0310-5.
58. Seki, M.; Narusaka, M.; Ishida, J.; Nanjo, T.; Fujita, M.; Oono, Y.; Kamiya, A.; Nakajima, M.; Enju, A.; Sakurai, T.; Satou, M.; Akiyama, K.; Taji, T.; Yamaguchi-Shinozaki, K.; Carninci, P.; Kawai, J.; Hayashizaki, Y.; Shinozaki, K. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray: Expression profiling under abiotic stresses. *Plant J.* **2002**, *31*, 279–292, doi:10.1046/j.1365-313X.2002.01359.x.

59. Kushiro, T.; Okamoto, M.; Nakabayashi, K.; Yamagishi, K.; Kitamura, S.; Asami, T.; Hirai, N.; Koshiha, T.; Kamiya, Y.; Nambara, E. The Arabidopsis cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism. *EMBO J.* **2004**, *23*, 1647–1656, doi:10.1038/sj.emboj.7600121.
60. Yokosho, K.; Yamaji, N.; Fujii-Kashino, M.; Ma, J. F. Functional analysis of a MATE gene OsFRDL2 revealed its involvement in Al-induced secretion of citrate, but a lower contribution to Al tolerance in rice. *Plant Cell Physiol.* **2016**, *57*, 976–985, doi:10.1093/pcp/pcw026.
61. Sun, C.; Zhang, F.; Yan, X.; Zhang, X.; Dong, Z.; Cui, D.; Chen, F. Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. *Plant Biotechnol. J.* **2017**, *15*, 953–969, doi:10.1111/pbi.12690.
62. Zanke, C. D.; Ling, J.; Plieske, J.; Kollers, S.; Ebmeyer, E.; Korzun, V.; Argillier, O.; Stiewe, G.; Hinze, M.; Neumann, K.; Ganai, M. W.; Röder, M. S. Whole *Genome Association Mapping of Plant Height in Winter Wheat* (*Triticum aestivum* L.). *PLoS ONE* **2014**, *9*, e113287, doi:10.1371/journal.pone.0113287.
63. Golabadi, M.; Arzani, A.; Mirmohammadi Maibody, S. A. M.; Sayed Tabatabaei, B. E.; Mohammadi, S. A. Identification of microsatellite markers linked with yield components under drought stress at terminal growth stages in durum wheat. *Euphytica* **2011**, *177*, 207–221, doi:10.1007/s10681-010-0242-8.

64. Yan, J.; Su, P.; Wei, Z.; Nevo, E.; Kong, L. Genome-wide identification, classification, evolutionary analysis and gene expression patterns of the protein kinase gene family in wheat and *Aegilops tauschii*. *Plant Mol. Biol.* **2017**, *95*, 227–242, doi:10.1007/s11103-017-0637-1.
65. Sakamoto, H.; Maruyama, K.; Sakuma, Y.; Meshi, T.; Iwabuchi, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Arabidopsis Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. *Plant Physiol.* **2004**, *136*, 2734–2746, doi:10.1104/pp.104.046599.
66. Ciftci-Yilmaz, S.; Morsy, M. R.; Song, L.; Coutu, A.; Krizek, B. A.; Lewis, M. W.; Warren, D.; Cushman, J.; Connolly, E. L.; Mittler, R. The ear-motif of the Cys2/His2-type zinc finger protein *Zat7* plays a key role in the defense response of arabidopsis to salinity stress. *J. Biol. Chem.* **2007**, *282*, 9260–9268, doi:10.1074/jbc.M611093200.
67. Ma, K.; Xiao, J.; Li, X.; Zhang, Q.; Lian, X. Sequence and expression analysis of the C3HC4-type RING finger gene family in rice. *Gene* **2009**, *444*, 33–45, doi:10.1016/j.gene.2009.05.018.
68. Cabral, A. L.; Jordan, M. C.; Larson, G.; Somers, D. J.; Humphreys, D. G.; McCartney, C. A. Relationship between QTL for grain shape, grain weight, test weight, milling yield, and plant height in the spring wheat cross RL4452/‘AC Domain.’ *PLOS ONE* **2018**, *13*, e0190681, doi:10.1371/journal.pone.0190681.
69. Zhou, J.; Wang, X.; Jiao, Y.; Qin, Y.; Liu, X.; He, K.; Chen, C.; Ma, L.; Wang, J.; Xiong, L.; Zhang, Q.; Fan, L.; Deng, X. W. Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. *Plant Mol. Biol.* **2007**, *63*, 591–608, doi:10.1007/s11103-006-9111-1.



70. Sourdille, P.; Cadalen, T.; Gay, G.; Gill, B.; Bernard, M. Molecular and physical mapping of genes affecting awning in wheat. *Plant Breed.* **2002**, *121*, 320–324, doi:10.1046/j.1439-0523.2002.728336.x.
71. Echeverry-Solarte, M.; Kumar, A.; Kianian, S.; Mantovani, E. E.; McClean, P. E.; Deckard, E. L.; Elias, E.; Simsek, S.; Alamri, M. S.; Hegstad, J.; Schatz, B.; Mergoum, M. Genome-wide mapping of spike-related and agronomic traits in a common wheat population derived from a supernumerary spikelet parent and an elite parent. *Plant Genome* **2015**, *8*, 0, doi:10.3835/plantgenome2014.12.0089.
72. Li, H.-Y.; Li, H.-Y.; Wang, T.-Y.; Li, H.-Y.; Wang, T.-Y.; Shi, Y.-S.; Li, H.-Y.; Wang, T.-Y.; Shi, Y.-S.; Fu, J.-J.; Li, H.-Y.; Wang, T.-Y.; Shi, Y.-S.; Fu, J.-J.; Song, Y.-C.; Li, H.-Y.; Wang, T.-Y.; Shi, Y.-S.; Fu, J.-J.; Song, Y.-C.; Wang, G.-Y.; Li, H.-Y.; Wang, T.-Y.; Shi, Y.-S.; Fu, J.-J.; Song, Y.-C.; Wang, G.-Y.; Li, Y. Isolation and characterization of induced genes under drought stress at the flowering stage in maize (*Zea mays*): Full Length Research Paper. *DNA Seq.* **2007**, *18*, 445–460, doi:10.1080/10425170701292051.
73. Pandey, S.; Assmann, S. M. The Arabidopsis putative g protein-coupled receptor GCR1 interacts with the G protein  $\alpha$  subunit GPA1 and regulates abscisic acid signaling. *Plant Cell* **2004**, *16*, 1616–1632, doi:10.1105/tpc.020321.
74. Wu, Q.; Chen, Y.; Fu, L.; Zhou, S.; Chen, J.; Zhao, X.; Zhang, D.; Ouyang, S.; Wang, Z.; Li, D.; Wang, G.; Zhang, D.; Yuan, C.; Wang, L.; You, M.; Han, J.; Liu, Z. QTL mapping of flag leaf traits in common wheat using an integrated high-density SSR and SNP genetic linkage map. *Euphytica* **2016**, *208*, 337–351, doi:10.1007/s10681-015-1603-0.

75. Yang, D.; Liu, Y.; Cheng, H.; Chang, L.; Chen, J.; Chai, S.; Li, M. Genetic dissection of flag leaf morphology in wheat (*Triticum aestivum* L.) under diverse water regimes. *BMC Genet.* **2016**, *17*, doi:10.1186/s12863-016-0399-9.
76. Liu, K.; Xu, H.; Liu, G.; Guan, P.; Zhou, X.; Peng, H.; Yao, Y.; Ni, Z.; Sun, Q.; Du, J. QTL mapping of flag leaf-related traits in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2018**, *131*, 839–849, doi:10.1007/s00122-017-3040-z.
77. Ros, B.; Thümmeler, F.; Wenzel, G. Analysis of differentially expressed genes in a susceptible and moderately resistant potato cultivar upon *Phytophthora infestans* infection. *Mol. Plant Pathol.* **2004**, *5*, 191–201, doi:10.1111/j.1364-3703.2004.00221.x.
78. Zhao, Y.; Cai, M.; Zhang, X.; Li, Y.; Zhang, J.; Zhao, H.; Kong, F.; Zheng, Y.; Qiu, F. Genome-wide identification, evolution and expression analysis of mTERF gene family in maize. *PLoS ONE* **2014**, *9*, e94126, doi:10.1371/journal.pone.0094126.
79. Cotsaftis, O.; Plett, D.; Johnson, A. A. T.; Walia, H.; Wilson, C.; Ismail, A. M.; Close, T. J.; Tester, M.; Baumann, U. Root-specific transcript profiling of contrasting rice genotypes in response to salinity stress. *Mol. Plant* **2011**, *4*, 25–41, doi:10.1093/mp/ssq056.
80. Hai, L.; Guo, H.; Xiao, S.; Jiang, G.; Zhang, X.; Yan, C.; Xin, Z.; Jia, J. Quantitative trait loci (QTL) of stem strength and related traits in a doubled-haploid population of wheat (*Triticum aestivum* L.). *Euphytica* **2005**, *141*, 1–9, doi:10.1007/s10681-005-4713-2.
81. Shumayla; Sharma, S.; Kumar, R.; Mendu, V.; Singh, K.; Upadhyay, S. K. Genomic dissection and expression profiling revealed functional divergence in *Triticum aestivum* leucine rich repeat receptor like kinases (TaLRRKs). *Front. Plant Sci.* **2016**, *7*, doi:10.3389/fpls.2016.01374.

82. Landjeva, S.; Neumann, K.; Lohwasser, U.; Börner, A. Molecular mapping of genomic regions associated with wheat seedling growth under osmotic stress. *Biol. Plant.* **2008**, *52*, 259–266, doi:10.1007/s10535-008-0056-x.
83. Ren, Y.; He, X.; Liu, D.; Li, J.; Zhao, X.; Li, B.; Tong, Y.; Zhang, A.; Li, Z. Major quantitative trait loci for seminal root morphology of wheat seedlings. *Mol. Breed.* **2012**.
84. Bai, C.; Liang, Y.; Hawkesford, M. J. Identification of QTLs associated with seedling root traits and their correlation with plant height in wheat. *J. Exp. Bot.* **2013**, *64*, 1745–1753, doi:10.1093/jxb/ert041.
85. Atkinson, J. A.; Wingen, L. U.; Griffiths, M.; Pound, M. P.; Gaju, O.; Foulkes, M. J.; Le Gouis, J.; Griffiths, S.; Bennett, M. J.; King, J.; Wells, D. M. Phenotyping pipeline reveals major seedling root growth QTL in hexaploid wheat. *J. Exp. Bot.* **2015**, *66*, 2283–2292, doi:10.1093/jxb/erv006.
86. Canè, M. A.; Maccaferri, M.; Nazemi, G.; Salvi, S.; Francia, R.; Colalongo, C.; Tuberosa, R. Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Mol. Breed.* **2014**, *34*, 1629–1645, doi:10.1007/s11032-014-0177-1.
87. Maccaferri, M.; El-Feki, W.; Nazemi, G.; Salvi, S.; Canè, M. A.; Colalongo, M. C.; Stefanelli, S.; Tuberosa, R. Prioritizing quantitative trait loci for root system architecture in tetraploid wheat. *J. Exp. Bot.* **2016**, *67*, 1161–1178, doi:10.1093/jxb/erw039.
88. Oono, Y.; Yazawa, T.; Kanamori, H.; Sasaki, H.; Mori, S.; Handa, H.; Matsumoto, T. Genome-Wide Transcriptome Analysis of Cadmium Stress in Rice. *BioMed Res. Int.* **2016**, *2016*, doi:10.1155/2016/9739505.

89. Furukawa, J.; Yamaji, N.; Wang, H.; Mitani, N.; Murata, Y.; Sato, K.; Katsuhara, M.; Takeda, K.; Ma, J. F. An Aluminum-Activated Citrate Transporter in Barley. *Plant Cell Physiol.* **2007**, *48*, 1081–1091, doi:10.1093/pcp/pcm091.
90. Ryan, P. R.; Raman, H.; Gupta, S.; Horst, W. J.; Delhaize, E. A second mechanism for aluminum resistance in wheat relies on the constitutive efflux of citrate from roots. *Plant Physiol.* **2009**, *149*, 340–351, doi:10.1104/pp.108.129155.
91. Baron, K. N.; Schroeder, D. F.; Stasolla, C. GEm-Related 5 (GER5), an ABA and stress-responsive GRAM domain protein regulating seed development and inflorescence architecture. *Plant Sci. Int. J. Exp. Plant Biol.* **2014**, *223*, 153–166, doi:10.1016/j.plantsci.2014.03.017.
92. Liu, L.; Li, N.; Yao, C.; Meng, S.; Song, C. Functional analysis of the ABA-responsive protein family in ABA and stress signal transduction in *Arabidopsis*. *Chin. Sci. Bull.* **2013**, *58*, 3721–3730, doi:10.1007/s11434-013-5941-9.
93. Singh, A.; Giri, J.; Kapoor, S.; Tyagi, A. K.; Pandey, G. K. Protein phosphatase complement in rice: genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development. *BMC Genomics* **2010**, *11*, 435, doi:10.1186/1471-2164-11-435.

## TABLES

Table 1. Mean monthly temperatures and total monthly rainfalls in two growing seasons (2016 and 2017) and 25-year average data in Konya, Turkey.

Month	Konya, 2015-2016 Temperature (temp) (°C) <sup>a</sup>	Konya, 2016-2017 Temp (°C)	Turkey, 1991-2015 Temp (°C) <sup>b</sup>	Konya, 2015-2016 Rainfall (mm) <sup>c</sup>	Konya, 2016- 2017 Rainfall (mm)	Turkey, 1991-2015 Rainfall (mm) <sup>d</sup>
September	22.8	17.1	19.0	35.8	11.2	23.1
October	15.3	13.2	13.7	34.4	0.0	48.3
November	7.5	4.9	7.0	5.8	16.6	58.0
December	-0.1	0.5	2.1	8.0	26.8	73.0
January	1.6	0.2	0.1	37.0	9.0	65.6
February	4.9	3.4	1.3	0.4	69.2	60.0
March	8.5	8.2	5.3	37.8	31.0	61.6
April	15.2	12.7	10.4	9.6	33.2	62.7
May	18.4	16.7	15.2	38.4	41.2	54.6
June	23.7	24.4	19.5	15.0	4.8	34.7
July	26.6	27.7	22.9	0.2	0.0	15.1
Total/average	13.1	11.7	10.6	222.4	243.0	435.1

<sup>a</sup>Source: Bahri Dagdas International Agricultural Research Institute

<sup>b</sup>Source: [http://sdwebx.worldbank.org/climateportal/index.cfm?page=country\\_historical\\_climate&ThisCCCode=TUR](http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisCCCode=TUR)

<sup>c</sup>Source: Bahri Dagdas International Agricultural Research Institute

<sup>d</sup>Source: [http://sdwebx.worldbank.org/climateportal/index.cfm?page=country\\_historical](http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical)

∞ **Table 2.** Phenotypic variation for grain yield and yield-related traits with best linear unbiased predictor values, range, percentage of coefficient of variation (CV), and broad sense heritability ( $H^2$ ) of 123 synthetic hexaploid wheat grown in two seasons (2016 and 2017) in Konya, Turkey.

Trait	2016				2017			
	Mean	Range	CV	$H^2$	Mean	Range	CV	$H^2$
Grain yield (gm <sup>-2</sup> )	259	200–341	9.7	0.32	290	241–392	9.9	0.56
Harvest index	0.4	0.24-0.66	10.9	0.63	0.34	0.27–0.41	6.3	0.64
Biomass weight (gm <sup>-2</sup> )	671	537–827	9.1	0.39	865	684–1098	8.9	0.63
Thousand kernel weight (g)	32.1	24–42	10.5	0.75	41	33–50	8	0.90
Grain volume weight (Kg hl <sup>-1</sup> )	65.6	52–77	7.2	0.91	74	68–77	2.3	0.76
Awn length (cm)	6	2.3–8.6	24.3	0.61	5.6	0.5–8.0	28.3	0.95
Flag leaf length (cm)	22.4	21.8–22.8	0.8	0.91	12	9.9–16.4	7.6	0.53
Flag leaf width (cm)	1	0.96–1.13	2.8	0.67	1	0.9–1.3	6.1	0.49
Flag leaf area (cm <sup>2</sup> )	18.9	17.6–19.7	2.2	0.85	10.1	7.7–14	11.6	0.52
Stem diameter (mm)	2.9	2.4–3.5	6.9	0.57	2.9	2.5–4.0	7.4	0.63
Root length (cm)	393	392–395	0.20	0.6	192.2	72–375	20	0.31

Table 3. List of significant markers associated with grain yield and yield-related traits, favorable alleles (underlined), SNP effects, and drought-related putative genes from genome-wide association study of 123 drought stressed synthetic hexaploid wheat grown in 2016 in Konya, Turkey.

Trait	SNP <sup>a</sup>	$-\log_{10}(P)$	Alleles	SNP Effect	PVE (%)	Gene-ID	Annotation
GY	S3A_686179591	4.08	<u>A</u> /G	-14.28	10.7	TraesCS3A01G445100	F-box family protein
GY	S7A_112977027	5.24	<u>A</u> /T	-19.17	12.8	TraesCS7A01G158200.1	Sentrin-specific protease
HI	S3A_593313534	13.56	<u>T</u> /C	0.08	16	TraesCS3A01G343700	WRKY transcription factor
HI	S6D_157451060	4.01	<u>A</u> /G	-0.03	6.2	TraesCS6D01G170900.1	Cytochrome P450, putative
HI	S6D_462272376	12.01	<u>G</u> /A	0.02	14.5	TraesCS6D01G382600.1	LOB domain protein-like
BMWT	S1D_441309135	4.82	<u>C</u> /G	105.34	14.4	TraesCS1D01G357500.1	Protein DETOXIFICATION
BMWT	S7B_450630784	4.06	<u>A</u> /G	-25.88	10.7	TraesCS7B01G242600.1	F-box family protein
TKW	S2A_47781717	4.52	<u>G</u> /A	0.93	4.2	TraesCS2A01G093500	F-box family protein
TKW	S4A_625466381	4.12	<u>T</u> /G	1.21	15.3	TraesCS4A01G347600	Protein kinase family protein
TKW	S4D_509427923	4.91	<u>C</u> /G	-1.72	10.1	TraesCS4D01G364700	Cytochrome P450 family protein
TKW	S6D_452410667	8.16	<u>A</u> /G	-1.54	17.7	TraesCS6D01G360800	Protein kinase family protein
AWNLN	S4D_461573496	5.71	<u>T</u> /C	0.32	9	TraesCS4D01G290700.1	60S ribosomal protein L18a
AWNLN	S5A_562540562	11.67	<u>C</u> /T	-1.71	11.3	TraesCS5A01G361300.1	Guanine nucleotide exchange family protein
FLLN	S1B_667135914	4.38	<u>C</u> /T	-0.16	20.8	TraesCS1B01G447400	Disease resistance protein RPM1
FLW	S6D_16376439	4.85	<u>C</u> /T	-0.02	13.3	TraesCS6D01G040100.1	Mitochondrial transcription termination factor-like
FLA	S1D_278097355	4.74	<u>G</u> /C	0.21	11.5	TraesCS1D01G197200.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
FLA	S6B_120860110	4.01	<u>G</u> /A	0.17	9.3	TraesCS6B01G125800	Cytochrome P450 family protein, expressed
FLA	S6B_120860130	4.01	<u>A</u> /T	-0.17	9.3	TraesCS6B01G125900	Cytochrome P450 family protein, expressed
STMDIA	S1D_431523575	6.58	<u>A</u> /G	-0.06	10.3	TraesCS1D01G341500	Disease resistance protein (NBS-LRR class) family
STMDIA	S3D_10133372	9.83	<u>G</u> /T	-0.11	8.6	TraesCS3D01G028500.1	Leucine-rich repeat receptor-like protein kinase family protein
STMDIA	S6A_94238211	6.9	<u>T</u> /G	0.06	7.5	TraesCS6A01G122200.1	Protein kinase, putative
RTLN	S5B_669373985	4.62	<u>T</u> /C	0.27	6.9	TraesCS5B01G502200	GRAM domain-containing protein / ABA-responsive
RTLN	S5B_669374027	4.62	<u>T</u> /C	0.27	6.9	TraesCS5B01G502200	GRAM domain-containing protein / ABA-responsive
RTLN	S6D_431108774	4.01	<u>A</u> /G	-0.27	5.8	TraesCS6D01G332800.1	Protein DETOXIFICATION
RTLN	S7A_94404310	4.01	<u>G</u> /A	0.52	7.5	TraesCS7A01G143200.2	Phosphatase 2C family protein

PVE: Phenotypic variance explained; GY, Grain yield; HI, Harvest index; BMWT, Biomass weight; TKW, Thousand kernel weight; GVWT, Grain volume weight; AWNLN, Awn length; FLLN, Flag leaf length; FLW, Flag leaf width; FLA, Flag leaf area; STMDIA, Stem diameter; RTLN, Root length. <sup>a</sup>S+chromosome\_chromosome position in bp.\_

Table 4. List of significant markers associated with grain yield and yield-related traits, favorable alleles (underlined), SNP effects, and drought-related putative genes obtained from genome-wide association study of 123 drought stressed synthetic hexaploid wheat grown in 2017 in Konya, Turkey.

Trait	SNP <sup>a</sup>	$-\log_{10}(P)$	Alleles	SNP Effect	PVE (%)	Gene-ID	Annotation
GY	S3A_25012018	4.81	<u>A</u> /G	-20.02	12.7	TraesCS3A01G047300	F-box domain containing protein
GY	S3D_1203058	4.12	<u>T</u> /G	14.32	12.8	TraesCS3D01G002700	Disease resistance protein RPM1
BMWT	S3A_25012018	6.08	<u>A</u> /G	-59.44	14.4	TraesCS3A01G047300	F-box domain containing protein
TKW	S4B_11905230	8.94	<u>C</u> /G	-1.11	3.9	TraesCS4B01G016200.1	LOB domain-containing protein, putative
TKW	S4B_637722874	5.17	<u>T</u> /C	0.86	2	TraesCS4B01G344200.1	Zinc finger (C3HC4-type RING finger) family protein
GVWT	S1A_522189599	4.11	<u>A</u> /G	-0.55	2.5	TraesCS1A01G334800	Cytochrome P450
GVWT	S4A_73454791	5.64	<u>C</u> /T	-0.63	5.5	TraesCS4A01G074200.2	Microtubule associated protein family protein, putative, expressed
AWNLN	S5B_43896804	7.31	<u>C</u> /T	-1.13	6	TraesCS5B01G038700	F-box family protein
AWNLN	S6B_643657	10.55	<u>C</u> /T	-1.04	5.7	TraesCS6B01G001000	F-box family protein
FLLN	S1B_631203243	5.26	<u>A</u> /G	-0.21	9.8	TraesCS1B01G400600.1	Rp1-like protein
FLLN	S2B_140752747	4.15	<u>G</u> /C	0.29	1.6	TraesCS2B01G167500.1	Cytochrome P450, putative
FLLN	S2D_642055122	4.12	<u>T</u> /C	0.25	5.9	TraesCS2D01G579800	protein kinase family protein
FLLN	S4A_612662321	5.3	<u>C</u> /T	-0.23	5.3	TraesCS4A01G325200	F-box family protein
FLLN	S6D_463762312	5.72	<u>G</u> /A	0.25	6	TraesCS6D01G386300	Cytochrome P450, putative
FLW	S1A_516732460	6.99	<u>A</u> /G	-0.03	9.5	TraesCS1A01G326700.1	Citrate-binding protein
FLW	S6B_26200560	7.13	<u>C</u> /A	0.03	12.3	TraesCS6B01G042800	F-box family protein
FLA	S1A_516732460	6.9	<u>A</u> /G	-0.42	8	TraesCS1A01G326700.1	Citrate-binding protein
FLA	S2A_764065400	4.18	<u>G</u> /T	-0.19	3.8	TraesCS2A01G563200	NBS-LRR resistance-like protein
STMDIA	S6B_610963076	5.7	<u>T</u> /G	0.06	7.9	TraesCS6B01G346900- TraesCS6B01G347000	NBS-LRR disease resistance protein and F-box protein-like
RTLN	S2D_620326979	4.22	<u>T</u> /C	192.21	9.9	TraesCS2D01G541000.1	Disease resistance protein RPM1

PVE: Phenotypic variance explained; GY, Grain yield; HI, BMWT, Biomass weight; TKW, Thousand kernel weight; GVWT, Grain volume weight; AWNLN, Awn length; FLLN, Flag leaf length; FLW, Flag leaf width; FLA, Flag leaf area; STMDIA, Stem diameter; RTLN, Root length. <sup>a</sup>S+chromosome\_chromosome position\_bp.



## FIGURES

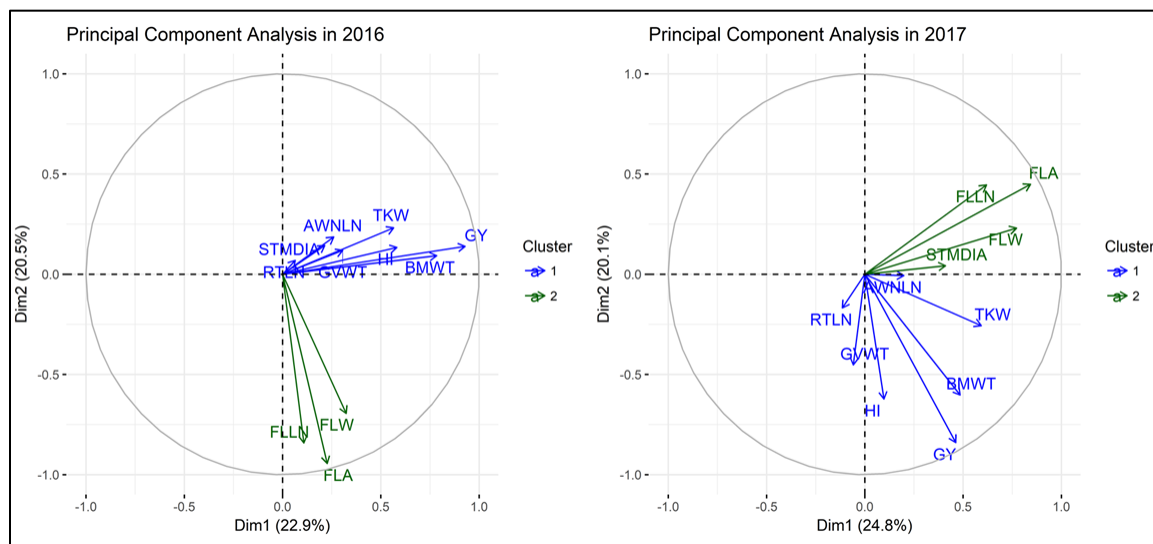


Figure 1. Principal component bi-plot analysis of 123 drought-stressed synthetic hexaploid wheat grown in two seasons (A: 2016 and B: 2017) in Konya, Turkey. AWLN, awn length; BMWT, biomass weight; FLA, flag leaf area; FLLN, flag leaf length; FLW, flag leaf width; GVWT, grain volume weight; GY, grain yield; HI, harvest index; RTLN, root length; STMDIA, stem diameter; and TKW, thousand kernel weight.

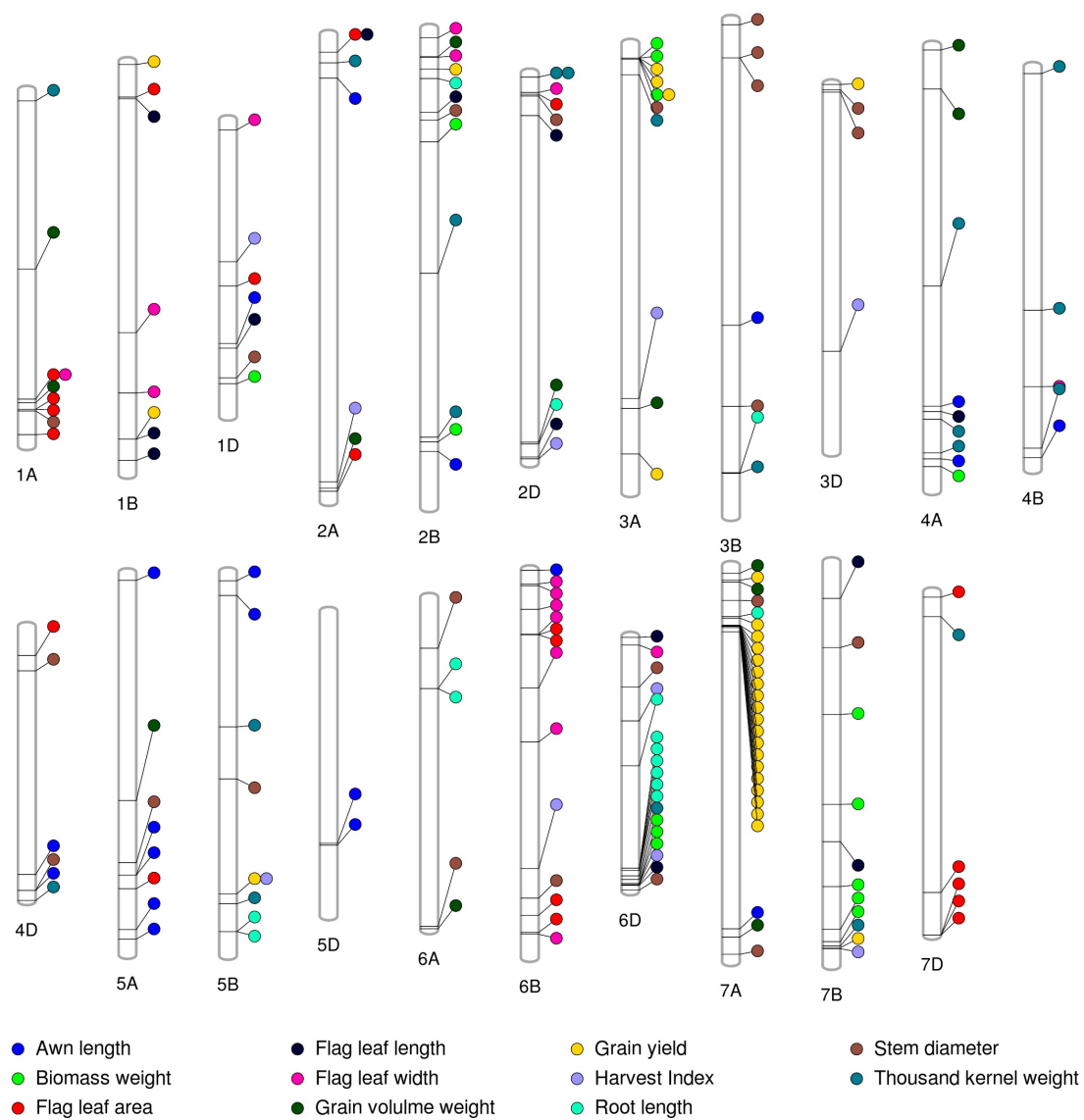


Figure 2. Significant marker trait associations identified on each chromosome for grain yield and yield-related traits obtained from genome-wide association study of 123 synthetic hexaploid wheat grown in 2016 and 2017 in Konya, Turkey.

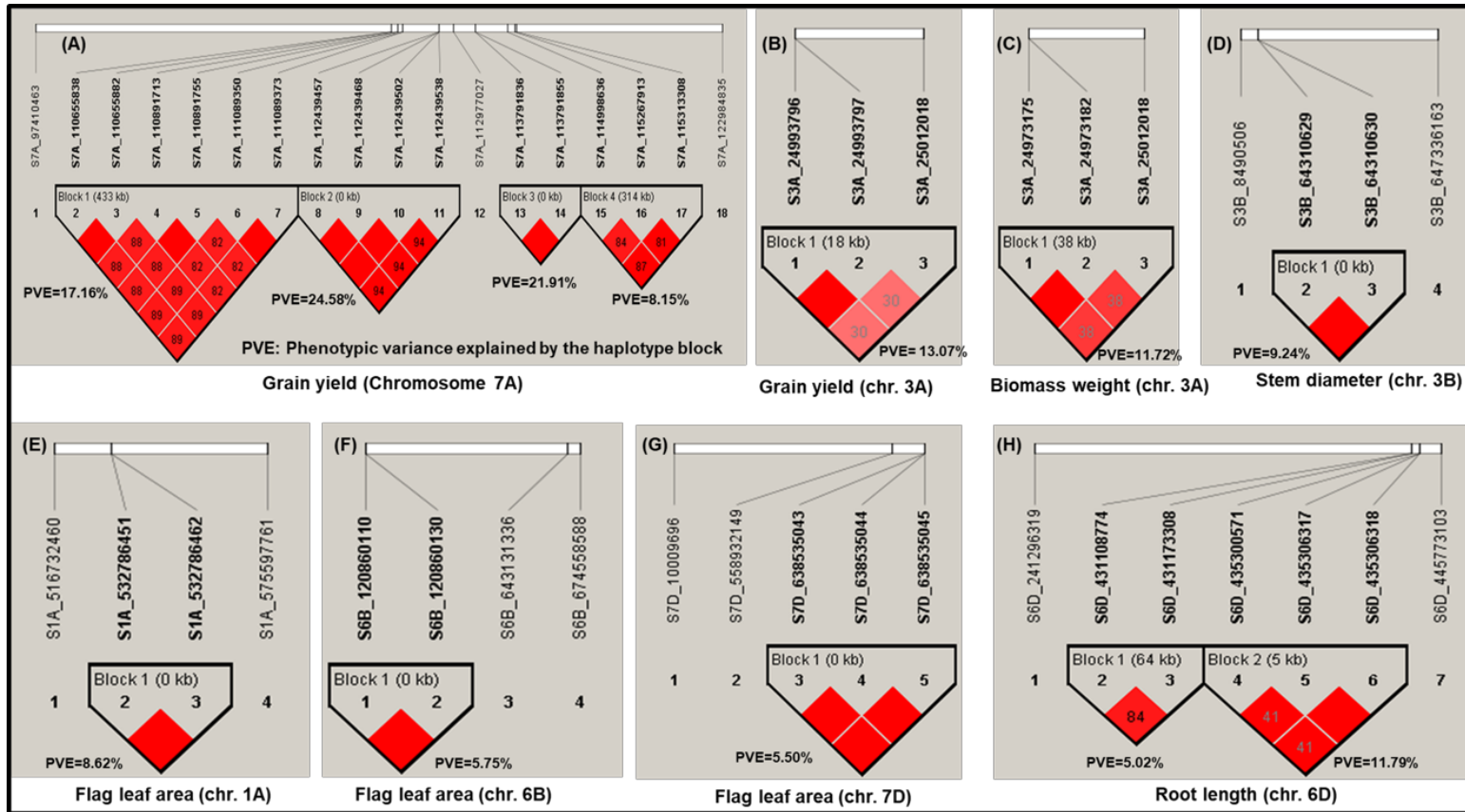
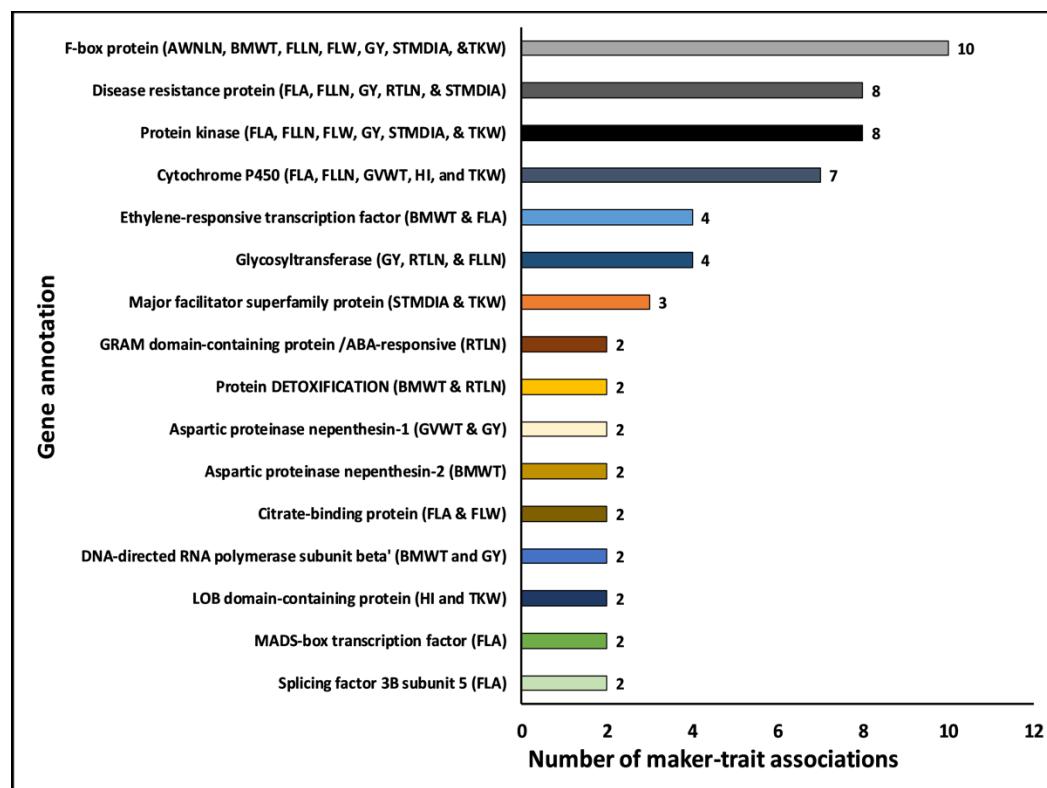


Figure 3. Linkage disequilibrium (LD) values ( $R^2$ ) and haplotype blocks with significant marker-trait associations (MTAs;  $\geq 2$ ) observed (A) on chromosome 7A for grain yield (B) on chromosome 3A for grain yield (C) on chromosome 3A for biomass weight (D) on chromosome 3B for stem diameter (E) on chromosome 1A for flag leaf area (F) on chromosome 6B for flag leaf area (G) on chromosome 7D for flag leaf area (H) on chromosome 6D for root length and phenotypic variance explained (PVE) by each haplotype block. Dark red color represents the strong LD whereas light red color represents the weak LD between pairs of MTAs.



**Figure 4.** Potential candidate gene functions harboring SNPs affecting yield and yield-related traits under drought stress. The count of marker-trait associations (for either single or multiple traits) located within genes that have the same gene annotation is shown. AWLN, awn length; BMWT, biomass weight; FLA, flag leaf area; FLLN, flag leaf length; FLW, flag leaf width; GVWT, grain volume weight; GY, grain yield; HI, harvest index; RTLN, root length; STMDIA, stem diameter; and TKW, thousand kernel weight.

## CHAPTER 4. GENOME-WIDE ASSOCIATION STUDY REVEALS FAVORABLE ALLELES ASSOCIATED WITH COMMON BUNT RESISTANCE IN SYNTHETIC HEXAPLOID WHEAT

This chapter is published: Bhatta M., Morgounov A., Belamkar V., Yorgancilar A., Baenziger P.S. *Euphytica* 214:200 (2018). <https://doi.org/10.1007/s10681-018-2282-4>

### Abstract

Genetic resistance to common bunt is a cost-effective, environmentally friendly, and sustainable approach to controlling the disease. To date, 16 race specific common bunt resistance genes (*Bt1-Bt15* and *Btp*) have been reported in wheat. However, a limited number have been mapped and few markers have been identified, which limits the usage of molecular markers in a marker-assisted breeding program. A total of 125 synthetic hexaploid wheats (SHWs) were evaluated for reactions to a mixture of common bunt races under field conditions in Turkey in 2016 and 2017. The objectives of this study were to identify common bunt resistant genotypes, identify genomic regions conferring resistance to common bunt using 35,798 genotyping-by-sequencing derived single nucleotide polymorphisms (SNPs), and investigate the significant SNPs present within genes using the functional annotations of the underlying genes. We found 29 resistant SHWs that can be used in wheat breeding. The genome-wide association study identified 15 SNPs associated with common bunt resistance and a haplotype block comprising three SNPs in perfect linkage disequilibrium. Five of them were novel and were located on chromosomes 2A, 3D, and 4A. Furthermore, seven of the 15 SNPs were found within genes and had annotations suggesting potential role in disease resistance. This study identified several favorable alleles that decreased common bunt incidence up to 26% in SHWs. These resistant SHWs and candidate genomic regions controlling common bunt resistance will be useful for

wheat genetic improvement and could assist in further understanding of the genetic architecture of common bunt resistance.

Keywords: Genetic resistance; Genomic regions; Genotyping-by-sequencing; Molecular markers; Resistance genes; Wheat genetic improvement

### **Abbreviations**

SHW, Synthetic Hexaploid Wheat; GBS, Genotyping-by-Sequencing; MTA, Marker Trait Association; BLUP, Best Linear Unbiased Predictor; GWAS, Genome Wide Association Study; MAF, Minor Allele Frequency; Q-Q plot, Quantile-Quantile plot; SNP, Single Nucleotide Polymorphism; IWGSC, International Wheat and Genome Sequencing Consortium

## Introduction

Common bunt [incited by *Tilletia tritici* (Bjerk.) and *T. laevis* (Kühn)] is one of the most important seed/soil-borne fungal diseases in wheat (*Triticum spp.*) (Wang et al. 2009; Goates 2012). Wheat yield and quality is reduced through the production of bunted sori filled with fungus teliospores (releasing fishy odor) in the grain (Wang et al. 2009). Although common bunt is effectively controlled through chemical seed treatment (Bhatta et al. 2018a), it increases the cost of production and may cause harm to the environment and human health (Wang et al. 2009). Chemical seed treatment with a synthetic pesticide is not allowed for controlling common bunt in organic farming (Fofana et al. 2008). Therefore, development and identification of genetic resistance to common bunt is the best method of controlling the disease in a cost-effective, environmentally friendly, and sustainable manner. Genetic pest resistance has been employed in agriculture for decades, however, effective genetic resistance against common bunt is still limited (He and Hughes 2003; Fofana et al. 2008).

To date, 16 race specific common bunt resistance genes (*Bt1-Bt15* and *Btp*) have been reported in wheat (Goates and Bockelman 2012; Goates 2012). However, only a limited number of genes have been mapped (Dumalasová et al. 2012), which is essential for marker-assisted breeding. Marker-assisted selection (MAS) can greatly facilitate common bunt resistance breeding as the phenotypic assay is difficult, time-consuming, environmentally dependent, and expensive. Additionally, MAS is useful for pyramiding resistance genes to provide durable resistance. The *Bt1* gene is located on chromosome 2B (Sears et al. 1960); *Bt4*, *Bt5*, and *Bt6* are on 1B (Schmidt et al. 1969; McIntosh 1998), *Bt7* on 2D (Schaller et al. 1960), *Bt9* on 6D (Steffan et al. 2017), *Bt10* on 6D (Menzies et al. 2006), and *Bt11* on 3B (Ciucă 2011). Quantitative trait loci (QTL) mapping of common bunt resistance identified 17 QTL on chromosomes 1B (Fofana

et al. 2008; Wang et al. 2009; Dumalasová et al. 2012; Galaev et al. 2012; Singh et al. 2016; Zou et al. 2017), 3A (Zou et al. 2017), 4B (Singh et al. 2016), 4D (Singh et al. 2016), 5B (Dumalasová et al. 2012; Singh et al. 2016), 7A (Fofana et al. 2008; Dumalasová et al. 2012), 7B (Dumalasová et al. 2012; Knox et al. 2013), and 7D (Singh et al. 2016). Of the 16 genes, *Bt10* is commonly used in wheat breeding programs due to its effectiveness against most of the races of common bunt (Menzies et al. 2006). New races are likely to evolve that will overcome previously identified (mostly race specific) resistance genes. For instance, Goates (2012) identified new races of common bunt virulent to *Bt8* or *Bt12*. Therefore, identifying and pyramiding new resistance genes are critical components of breeding programs that wish to achieve durable resistance.

One approach to identifying novel disease resistance genes is through the utilization of primary gene pool wild relatives. Synthetic hexaploid wheat (SHWs;  $2n=6x=42$ , AABBDD) (Bhatta et al. 2018b) is produced through hybridization of cultivated durum wheat ( $2n=4x=28$ , AABB, *Triticum durum* L.) and wild goat grass [ $(2n=2x=14, DD)$ , *Aegilops tauschii* Coss.]. The SHWs have been reported to have resistance to multiple abiotic (Bhatta et al. 2018c) and biotic stresses including common bunt (Morgounov et al. 2018). The SHWs under study have a large amount of genetic diversity especially for the D-genome compared to the elite bread wheat cultivars (Bhatta et al. 2018b). A genome-wide association study (GWAS) is useful for the identification of genomic regions associated with common bunt resistance. The objectives of this study were (i) to identify common bunt resistant genotypes, (ii) determine genomic regions associated with common bunt resistance among 125 SHWs using 35,798 genotyping-by-sequencing (GBS) derived single nucleotide polymorphisms (SNPs), and (iii) further investigate the significant SNPs present within genes using the functional annotations of the underlying



genes. This is the first report on GWAS for common bunt resistance in wheat and SHW. The results of this study will be valuable for accelerating the successful deployment of favorable alleles conferring resistance to common bunt through genomic and marker-assisted selection and for utilizing common bunt resistance from SHWs for broadening the genetic base of elite wheat germplasm.

## **Materials and Methods**

### Site description and plant materials

Field experiments were conducted in 2016 and 2017 at a research farm located in the Transitional Zone Agricultural Research Institute, Eskisehir, Turkey. The study population consisted of 125 SHW lines derived from crosses of seven durum parents crossed with 25 different *Ae. tauschii* accessions (APPENDIX I) and was described previously (Bhatta et al. 2018b). These SHW lines are available from the International Winter Wheat Improvement Program (<http://www.iwwip.org>) at CIMMYT in Turkey (Morgounov et al. 2018).

### Disease evaluation and experimental design

Phenotypic disease evaluation was performed under field conditions in the 2016 and 2017 growing seasons. Seeds were artificially inoculated with a mixture of spores (obtained from sampling naturally infected spikes) of different races of *T. tritici* and *T. laevis* prevalent in Eskisehir, Turkey. Seeds were then sown in one-meter row plot around mid-October in an augmented design with replicated checks ('Gerek' and 'Karahani') in both years. Disease incidence was estimated as a percentage of bunted spikes per row at plant maturity (around mid-July). The SHWs were determined as being resistant or susceptible to common bunt based on the

percentage of infected spikes: 0.0%= very resistant, 0.1 to 5.0%= resistant, 5.1 to 10.0%= moderately resistant, 10.1 to 30.0%= moderately susceptible, 30.1 to 50.0%= susceptible, and 50.1 to 100.0%= very susceptible (Veisz et al. 2000).

### Phenotypic data analysis

A combined (two-years) analysis of variance of the percentage of common bunt infected kernels was done assuming a mixed linear model. The PROC MIXED routine in SAS 9.4 (SAS Institute, 2018) was used to estimate the best linear unbiased predictors (BLUPs) and to determine the effects of genotype, year, and their interactions on common bunt resistance reactions using the model:  $y_{ijklm} = \mu + Yr_i + B(Yr)_{ji} + C_k + G_{l(ji)} + GxYr_{li} + e_{ijklm}$ , where,  $y_{ijklm}$  is the percentage of common bunt spikes;  $\mu$  is the overall mean;  $Yr_i$  is the effect of the  $i^{\text{th}}$  year;  $B(Yr)_{ji}$  is the effect of the  $j^{\text{th}}$  incomplete block within the  $i^{\text{th}}$  year;  $C_k$  is the  $k^{\text{th}}$  check;  $G_{l(ji)}$  is the effect of the  $l^{\text{th}}$  genotype (new variable, where check was coded as 0 and entry was coded as 1 and genotype was taken as new variable x entry) within  $j^{\text{th}}$  incomplete block of  $i^{\text{th}}$  year;  $GxYr_{li}$  is the interaction effect of  $l^{\text{th}}$  genotype and  $i^{\text{th}}$  year;  $e_{ijklm}$  is the residual. In this model, year and check were assumed to be fixed effects, whereas genotypes, genotype x year interaction, and incomplete blocks nested within a year were assumed to be random effects. Broad-sense heritability was calculated based on entry mean basis as described previously (Bhatta et al. 2018c). A Pearson correlation between the two years' datasets was calculated using PROC CORR in SAS using BLUPs.

## Genotyping and SNP Discovery

DNA extraction using BioSprint 96 DNA Plant Kits (Qiagen, Hombrechtikon, Switzerland), genotyping-by-sequencing, and SNP discovery were carried out as described in Bhatta et al. (2018b). SNP discovery was done using TASSEL v. 5.2.40 GBS v2 Pipeline (Glaubitz et al. 2014) with physical alignment to the Chinese Spring genome sequence (RefSeq v1.0) provided by the International Wheat Genome Sequencing Consortium (IWGSC 2018). After SNP discovery, the SNPs with minor allele frequency (MAF) less than 5% and more than 20% missing data were removed from the analysis Bhatta et al. (2018b, c). All lines had less than 20% missing sites and thus none were dropped from the analysis.

## Population structure, genome-wide association analysis, and candidate gene analysis

The population structure of the 125 genotypes was assessed using the Bayesian clustering algorithm in the program STRUCTURE v 2.3.4 (Pritchard et al. 2000) and principal component analysis (PCA) was done using TASSEL (Bradbury et al. 2007, as described in Bhatta et al. (2018b, c). In brief, three sub-populations were observed both in the STRUCTURE and PCA, and hence, the first three PCs were used in the GWAS to control population structure.

A GWAS was performed on the BLUPs combined over years to identify SNPs associated with common bunt resistance (marker-trait associations: MTAs) using FarmCPU (Fixed and random model Circulating Probability Unification) with the first three principal components (PC<sub>1</sub>, PC<sub>2</sub>, and PC<sub>3</sub>) and FarmCPU calculated kinship (Liu et al. 2016) implemented in MVP R software package (<https://github.com/XiaoleiLiuBio/MVP>). The identified MTAs were tested against a uniform suggestive genome-wide significance threshold ( $P = 8.8E-05$ ;  $-\log_{10}(P)$  value = 4.05) considering the deviation of Q-Q plot (Bhatta et al. 2018c). The percentage of phenotypic

variance explained ( $R^2$ ) by each SNP was calculated using TASSEL (Bradbury et al. 2007). Haplotype block analysis with linkage disequilibrium (LD) values in adjacent regions (<500 kb) of significant MTAs were visualized and plotted in a Haploview software using default parameters (Barret et al. 2005). To further understand the role of significant MTAs, functional annotations of the underlying genes containing these SNPs were retrieved from the IWGSC RefSeq v1.0 functional gene annotation provided for every gene in the genome (IWGSC 2018). The T-tests were performed to determine whether the favorable alleles significantly improved the trait of interest compared to the unfavorable alleles. Regression analysis was performed between the number of favorable alleles and common bunt incidence (%) to understand whether the favorable alleles significantly decreases the common bunt incidence in SHWs and vice-versa.

## Results and Discussion

### Phenotypic distribution of common bunt incidence in the SHW population

A combined analysis of variance of common bunt incidence showed the non-significant interaction between genotypes and years (mean squares = 66.6,  $P=0.99$ ) whereas the main effects of genotypes (mean squares =682.7,  $P<0.0001$ ) and years (mean squares = 632.1,  $P<0.0001$ ) were both significant. The non-significant genotype by year interaction implies that genotypes responded similarly across years. This result was also reflected by the strong correlation ( $r=0.95$ ,  $P<0.0001$ ) observed between genotypes across the two years. Therefore, the BLUPs obtained from the combined analysis of variance assuming a mixed model was used for further analysis. The present study identified moderate heritability ( $H^2=0.58$ ) for common bunt incidence. Low to moderate heritability for common bunt incidence was observed previously (Fofana et al. 2008).

The percentage of common bunt infection in SHWs based on the combined BLUPs ranged from 0.2 to 75.4% with an average of 20.8% (Fig. 1), indicating considerable genetic variation among SHWs. As expected, the bunt susceptible check cultivar, Gerek, showed a high (60.3%) disease incidence, whereas the bunt resistant check cultivar, Karahan, showed a low incidence (1.7%) (Fig. 1). Hence, our field assay was very effective in discriminating among the entries. Four of the SHWs were very resistant (<0.1% disease incidence) and 25 were resistant (0.1 to 5% disease incidence) to common bunt (Fig. 1). Twenty-one (~72%) resistant SHWs derived from Ukrainian durum parents ('UKR-OD 1530.94, LEUC 84693', and 'UKR-OD 761.93, AISBERG'), six of them (~21%) were from a Romanian durum (PANDUR) parent, and two of them (~7%) were from spring durum (Langdon) parent. The 29 resistant SHWs can be used in breeding programs for the introduction of common bunt resistance.

#### Genome-wide association study

A GWAS was performed to identify MTAs controlling resistance to common bunt using a multi-locus mixed model (MLMM) implemented in FarmCPU, where the combined BLUPs over two years (2016 and 2017) were the phenotypic data and 35,798 GBS-derived SNPs (MAF>5%, missing data <20%) were the genotypic data. The SNPs used in this study were well distributed across the 21 chromosomes in the SHWs (Fig. 2).

Genome-wide association study detected 15 MTAs for common bunt resistance on chromosomes 1B, 2A, 2B, 3D, 4A, 7A, and 7B (Fig. 3) with phenotypic variance explained (PVE) ranging from 1.22 to 16.7% (Table 1). Earlier studies have reported QTL for common bunt resistance on chromosomes 1B (Schmidt et al. 1969; McIntosh 1998; Fofana et al. 2008; Wang et al. 2009; Dumalasová et al. 2012; Galaev et al. 2012; Singh et al. 2016; Zou et al.

2017), 2B (Sears et al. 1960), 2D (Schaller et al. 1960), 3A (Zou et al. 2017), 3B (Ciucă 2011), 4B (Singh et al. 2016), 4D (Singh et al. 2016), 5B (Dumalasová et al. 2012; Singh et al. 2016), 6D (Menzies et al. 2006; Steffan et al. 2017), 7A (Fofana et al. 2008; Dumalasová et al. 2012), 7B (Dumalasová et al. 2012; Knox et al. 2013), and 7D (Singh et al. 2016). The five MTAs identified in this study on chromosomes 2A (three MTAs), 3D (one), and 4A (one) have not been previously reported and they are potentially novel MTAs. The present study identified a haplotype block (perfect LD; squared correlation coefficient between locus allele frequency i.e.,  $r^2 = 1$ ) of size 18 bp on chromosome 7A with three SNPs associated with common bunt resistance and the PVE by each MTA present on this haplotype block was 5.1%. Identification of a haplotype block with multiple MTAs on chromosome 7A as earlier studies (Fofana et al. 2008; Dumalasová et al. 2012) provided increased confidence on these associations and showed that the chromosome 7A is important for common bunt resistance in wheat.

Potential candidate genes underlying marker-trait associations and their gene annotations

The seven MTAs detected on chromosomes 2A (1), 2B (1), 4A (1), 7A (3), and 7B (1) were within genes whose annotation suggested that they could be associated with disease resistance (IWGSC 2018) (Table 1). For instance, the genes flanking SNP-S2A\_690962587 on chromosome 2A were annotated as clathrin interactor *EPSIN 1* (TraesCS2A01G440100.1) and ubiquinol oxidase/alternative oxidase (TraesCS2A01G438200.1). Alternative oxidase (PF01786) is a major stress-induced protein that has been found to be involved in maintaining metabolic and signaling homeostasis during abiotic and biotic stress in plants (Vanlerberghe 2013; Moore et al. 2013). The gene containing SNP-S2B\_799254305 on chromosome 2B was annotated as *CsAtPR5* (TraesCS2B01G627300.1). This gene is a precursor of the pathogenesis-related protein

5 (PR5) and has been found to be associated with yellow rust (incited by *Puccinia striiformis* f.sp. *tritici* Westend.) resistance (Bozkurt et al. 2007) and powdery mildew (incited by *Blumeria graminis* (DC.) Speer f.sp. *tritici*) resistance (Niu et al. 2010) in wheat. The genes flanking SNP-S4A\_721406696 on chromosome 4A were annotated as diacylglycerol O-acyltransferase 2 (TraesCS4A01G456400.2) and cellulose synthase-like protein or RING/Ubox like zinc-binding domain (TraesCS4A01G456300.1) and were previously found to be involved in biotic (Chen et al., 2016) and abiotic stress (salt and drought) tolerance (Zhu et al. 2010; Im et al. 2017). All three MTAs (SNP S7A\_703054426, S7A\_703054429, and S7A\_703054444) on chromosome 7A were within the same gene (TraesCS7A01G518800.1), which was annotated as NF-X1-type zinc finger protein NFXL1 (TraesCS7A01G518800.1) (Table 1). This gene has been reported to be involved in rice blast disease (incited by *Magnaporthe oryzae* Cavara) resistance (Li et al. 2014). The gene associated with SNP S7B\_660640575 on chromosome 7B were ENTH/ANTH/VHS superfamily protein (TraesCS7B01G393800.2) and Zinc finger protein such as PHD (plant homeodomain) finger (TraesCS7B01G393900.1). The PHD finger is a nuclear protein involved in regulating defense responses against abiotic and biotic stresses (Hong et al. 2007; Cheung et al. 2007). While these genes may have a role in controlling resistance to common bunt in the SHWs, the present data do not provide proof of this. However, the MTAs that were identified in this study provide evidence for evaluating these genes in future functional characterization studies for common bunt in wheat. Additionally, these MTAs can be explored in marker-assisted breeding.

### Favorable alleles associated with common bunt resistance

In the present study, alleles of MTAs that decrease the incidence of common bunt were considered favorable whereas alleles that increase the common bunt incidence were considered unfavorable. Allelic variation for significant MTAs in the 125 genotypes is presented in box plots (Fig. 4). The distribution of favorable/unfavorable alleles varied widely among SHWs (Fig. 4 and APPENDIX VIII) and the numbers of favorable alleles ranged from 0 to 11 with a median of 7 alleles (Fig. 4). The resistant genotypes (29) had favorable alleles ranging from 3 to 11 with a median of 8 alleles. The favorable alleles involved in MTAs decreased the percentage of common bunt infected spikes by 0.47% (S7A\_9996936\_G) up to 26 % (S2B\_799254305\_G) (Fig. 4). Six highly favorable alleles (G of SNP 2A\_690962587, G of SNP S2B\_795282008, G of SNP S2B\_799254305, T of SNP S3D\_575937403, T of SNP S4A\_721406696, and T of SNP S7B\_660640575) significantly ( $P < 0.05$ ) decreased common bunt incidence compared to the genotypes with unfavorable alleles (Fig. 4), indicating their potential usefulness in breeding programs. Regression analysis confirmed that common bunt incidence (%) linearly decreased with an increase in the number of favorable alleles, whereas it increased with an increase in the number of unfavorable alleles (Fig. 5). The PVE ( $R^2$ ) by favorable alleles on common bunt incidence was 19% whereas the PVE by unfavorable alleles was 25% (**Fig. 5**). These results suggested that the highly favorable alleles for common bunt resistance from SHWs could be used in elite wheat breeding program for pyramiding superior alleles.

### Conclusions

The present study on SHWs found a considerable genetic variation for common bunt incidence and identified 29 resistant SHWs that can be used in wheat breeding programs. A



marker-trait analysis was conducted using GWAS and identified 15 MTAs for common bunt resistance. Seven of the MTAs were within genes with functional annotations that suggest their involvement in disease resistance. This study also identified five novel MTAs, and of these, three were found within genes with disease resistance functions, which provided the confidence on the usefulness of the identified MTAs. Additionally, we identified three MTAs on a haplotype block in a perfect LD and the MTAs present in each haplotype block explained 5.1% of the phenotypic variance. We identified highly favorable alleles associated with common bunt resistance and the resistant SHWs identified in this study should be valuable for the introgression of resistance into elite bread wheat germplasm. Apart from their utility in the marker-aided selection, the markers identified here will facilitate a better understanding of the genetic basis of common bunt resistance.

## References

- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265. doi: 10.1093/bioinformatics/bth457
- Bhatta M, Regassa T (2018) Seed treatment affected yield and economic returns of Nebraska winter wheat genotypes. *J Agri Sci* 9:5-10. doi:11.258359/KRE-146
- Bhatta M, Morgounov A, Belamkar V, et al (2018b) Unlocking the novel genetic diversity and population structure of synthetic hexaploid wheat. *BMC Genomics* 19:951. doi.org/10.1186/s12864-018-4969-2
- Bhatta M, Morgounov A, Belamkar V, Baenziger PS (2018c) Genome-wide association study for grain yield and yield-related traits in drought-stressed synthetic hexaploid wheat germplasm. *Int J Mol Sci* 19:3011. doi.org/10.3390/ijms19103011
- Bozkurt O, Unver T, Akkaya MS (2007) Genes associated with resistance to wheat yellow rust disease identified by differential display analysis. *Physiol Mol Plant Pathol* 71:251–259. doi: 10.1016/j.pmpp.2008.03.002
- Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi: 10.1093/bioinformatics/btm308
- Chen B, Wang J, Zhang G, Liu J, Manan S, Hu H, Zhao J (2016). Two types of soybean diacylglycerol acyltransferases are differentially involved in triacylglycerol biosynthesis and response to environmental stresses and hormones. *Scientific reports* 6:28541. doi: 10.1038/srep28541

Cheung M-Y, Zeng N-Y, Tong S-W, et al (2007) Expression of a RING-HC protein from rice improves resistance to *Pseudomonas syringae* pv. tomato DC3000 in transgenic *Arabidopsis thaliana*. *J Exp Bot* 58:4147–4159. doi: 10.1093/jxb/erm272

Ciucă M (2011) A preliminary report on the identification of SSR markers for bunt (*Tilletia* sp.) resistance in wheat. *Czech J Genet Plant Breed* 47:S142–S145. doi: 10.17221/3269-CJGPB

DumalasoVá V, Simmonds J, Bartoš P, Snape J (2012) Location of genes for common bunt resistance in the European winter wheat cv. Trintella. *Euphytica* 186:257–264. doi: 10.1007/s10681-012-0671-7

Fofana B, Humphreys DG, Cloutier S, et al (2008) Mapping quantitative trait loci controlling common bunt resistance in a doubled haploid population derived from the spring wheat cross RL4452 × AC Domain. *Mol Breed* 21:317–325. doi: 10.1007/s11032-007-9131-9

Galaev AV, Babayants LT, Sivolap YM (2012) DNA-markers for resistance to common bunt transferred from *Aegilops cylindrica* Host. to hexaploid wheat. *Czech J Genet Plant Breed* 42:62–65. doi: 10.17221/6234-CJGPB

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLOS ONE* 9:e90346. doi: 10.1371/journal.pone.0090346

Goates BJ (2012) Identification of new pathogenic races of common bunt and dwarf bunt fungi, and evaluation of known races using an expanded set of differential wheat lines. *Plant Dis* 96:361–369. doi: 10.1094/PDIS-04-11-0339

Goates BJ, Bockelman HE (2012) Identification of New Sources of High Levels of Resistance to Dwarf Bunt and Common Bunt among Winter Wheat Landraces in the USDA-ARS National Small Grains Collection. *Crop Sci* 52:2595. doi: 10.2135/cropsci2012.01.0060

He and G. R. Hughes C (2003) Inheritance of resistance to common bunt in spelt and common wheat. *Can J Plant Sci* 83:47–56. doi: 10.4141/P01-167

Hong JK, Choi HW, Hwang IS, Hwang BK (2007) Role of a novel pathogen-induced pepper C3–H–C4 type RING-finger protein gene, *CaRFP1*, in disease susceptibility and osmotic stress tolerance. *Plant Mol Biol* 63:571–588. doi: 10.1007/s11103-006-9110-2

Im S, Lee H-N, Jung HS, et al (2017) Transcriptome-based identification of the desiccation response genes in marine red algae *Pyropia tenera* (Rhodophyta) and enhancement of abiotic stress tolerance by PtDRG2 in chlamydomonas. *Mar Biotechnol* 19:232–245. doi: 10.1007/s10126-017-9744-x

International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361(6403):aar7191. doi: 10.1126/science.aar7191

Knox RE, Campbell HL, Depauw RM, et al (2013) DNA markers for resistance to common bunt in ‘McKenzie’ wheat. *Can J Plant Pathol* 35:328–337. doi: 10.1080/07060661.2013.763292

Li W-T, Chen W-L, Yang C, et al (2014) Identification and network construction of zinc finger protein (*ZFP*) genes involved in the rice-*Magnaporthe oryzae* interaction. *Plant Omics* 7:540

- Liu X, Huang M, Fan B, et al (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genet* 12:e1005767. doi: 10.1371/journal.pgen.1005767
- McIntosh RA (1998) Catalogue of gene symbols for wheat. *Proc 9th Inter Wheat Genet Symp Univ Sask Saskat Can* 1998 5:119–120
- Menzies JG, Knox RE, Popovic Z, Procnier JD (2006) Common bunt resistance gene *Bt10* located on wheat chromosome 6D. *Can J Plant Sci* 86:1409–1412. doi: 10.4141/P06-106
- Moore AL, Shiba T, Young L, et al (2013) Unraveling the heater: New insights into the structure of the alternative oxidase. *Annu Rev Plant Biol* 64:637–663. doi: 10.1146/annurev-arplant-042811-105432
- Morgounov A, Abugalieva A, Akan K, et al (2018) High-yielding winter synthetic hexaploid wheats resistant to multiple diseases and pests. *Plant Genet Resour* 16:273–278. doi: 10.1017/S147926211700017X
- Niu J, Jia H, Yin J, et al (2010) Development of an STS marker linked to powdery mildew resistance genes *PmLK906* and *Pm4a* by gene chip hybridization. *Agric Sci China* 9:331–336. doi: 10.1016/S1671-2927(09)60101-2
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- SAS Institute. 2018. SAS 9.4 product documentation. SAS Inst., Cary, NC.  
<http://support.sas.com/documentation/94/index.html> (accessed 10 May. 2018)

Schaller CW, Holton CS, Kendrick EL (1960) Inheritance of the second factor for resistance to bunt, *Tilletia caries* and *T. foetida*, in the wheat variety Martin. *Agron J* 52:280–85

Scmidt JW, Morris R, Johnson VA (1969) Monosomic analysis for bunt resistance in derivatives of Turkey and oro wheats 1. *Crop Sci* 9:286–288. doi: 10.2135/cropsci1969.0011183X000900030009x

Sears ER, Schaller CW, Briggs FN (1960) Identification of the Chromosome Carrying the Martin Gene for Resistance of Wheat to Bunt. *Can J Genet Cytol* 2:262–267. doi: 10.1139/g60-026

Singh A, Knox RE, DePauw RM, et al (2016) Genetic mapping of common bunt resistance and plant height QTL in wheat. *Theor Appl Genet* 129:243–256. doi: 10.1007/s00122-015-2624-8

Steffan PM, Torp AM, Borgen A, et al (2017) Mapping of common bunt resistance gene Bt9 in wheat. *Theor Appl Genet* 130:1031–1040. doi: 10.1007/s00122-017-2868-6

Vanlerberghe G (2013) Alternative Oxidase: A Mitochondrial Respiratory Pathway to Maintain Metabolic and Signaling Homeostasis during Abiotic and Biotic Stress in Plants. *Int J Mol Sci* 14:6805–6847. doi: 10.3390/ijms14046805

Veisz O, Szunics L, Szunics L (2000) Effect of common bunt on the frost resistance and winter hardiness of wheat (*Triticum aestivum* L.) lines containing Bt genes. *Euphytica* 114:159–164. doi: 10.1023/A:1003940808041

Wang S, Knox RE, DePauw RM, et al (2009) Markers to a common bunt resistance gene derived from ‘Blizzard’ wheat (*Triticum aestivum* L.) and mapped to chromosome arm 1BS. *Theor Appl Genet* 119:541–553. doi: 10.1007/s00122-009-1063-9

Zhu J, Lee B-H, Dellinger M, et al (2010) A cellulose synthase-like protein is required for osmotic stress tolerance in Arabidopsis: SOS6 is important for osmotic stress tolerance in plants.

Plant J 63:128-140. doi: 10.1111/j.1365-313X.2010.04227.x

Zou J, Semagn K, Chen H, et al (2017) Mapping of QTLs associated with resistance to common bunt, tan spot, leaf rust, and stripe rust in a spring wheat population. Mol Breed 37:144. doi:

10.1007/s11032-017-0746-1

## TABLES

Table 1. List of significant markers associated with common bunt resistance, favorable allele (underlined), SNP effects, and functional annotation of genes containing or flanking significant SNPs from the genome-wide association study of 125 synthetic hexaploid wheats grown in 2016 and 2017 in Eskisehir, Turkey.

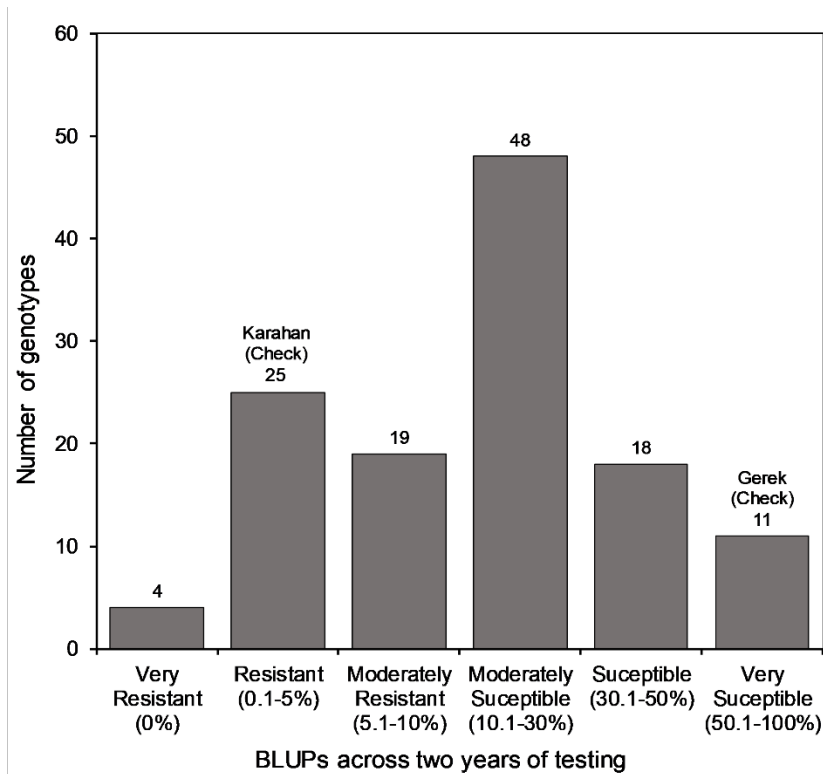
SNP <sup>a</sup>	- log <sub>10</sub> (P)	Alleles	SNP Effect	PVE (%) <sup>b</sup>	Gene-ID	Annotation
S1B_605788510	7.04	<u>G</u> /T	8.47	4.57	-	-
S2A_393513188	5.80	<u>C</u> /T	3.72	1.22	-	-
S2A_686298452	4.69	<u>C</u> /T	7.86	6.78	-	-
S2A_690962587	10.57	<u>G</u> /A	-6.86	8.01	TraesCS2A01G440100.1- TraesCS2A01G440200.1	Clathrin interactor EPSIN 1 - Ubiquinol oxidase/Alternative oxidase
S2B_795282008	11.74	<u>G</u> /C	12.29	15.34	-	-
S2B_799254305	12.94	<u>G</u> /C	-7.14	16.7	-	-
S3D_575937403	6.54	<u>T</u> /G	-5.57	4.01	TraesCS2B01G627300.1	<i>CsAtPR5</i>
S4A_721406696	4.55	<u>T</u> /C	-4.37	3.47	TraesCS4A01G456300.1- TraesCS4A01G456400.2	Cellulose synthase-like protein/RING/Ubox like zinc- binding domain - Diacylglycerol O-acyltransferase 2
S7A_41818214	4.06	<u>A</u> /G	5.07	10.84	-	-
S7A_703054426	4.25	<u>T</u> /C	-4.05	5.11	TraesCS7A01G518800.1- TraesCS7A01G518900.1	NF-X1-type zinc finger protein NFXL1 -APOLLO
S7A_703054429	4.25	<u>G</u> /C	-4.05	5.11	TraesCS7A01G518800.1- TraesCS7A01G518900.1	NF-X1-type zinc finger protein NFXL1 - APOLLO
S7A_703054444	4.25	<u>C</u> /T	4.05	5.11	TraesCS7A01G518800.1- TraesCS7A01G518900.1	NF-X1-type zinc finger protein NFXL1 -APOLLO
S7A_9996936	5.72	<u>G</u> /T	6.56	5.25	-	-
S7B_629415715	5.86	<u>A</u> /G	6.77	4.88	-	-
S7B_660640575	4.17	<u>C</u> /T	5.28	4.32	TraesCS7B01G393800.2- TraesCS7B01G393900.3	ENTH/ANTH/VHS superfamily protein -Zinc finger-PHD-finger

<sup>a</sup>S+chromosome\_chromosome position in bp

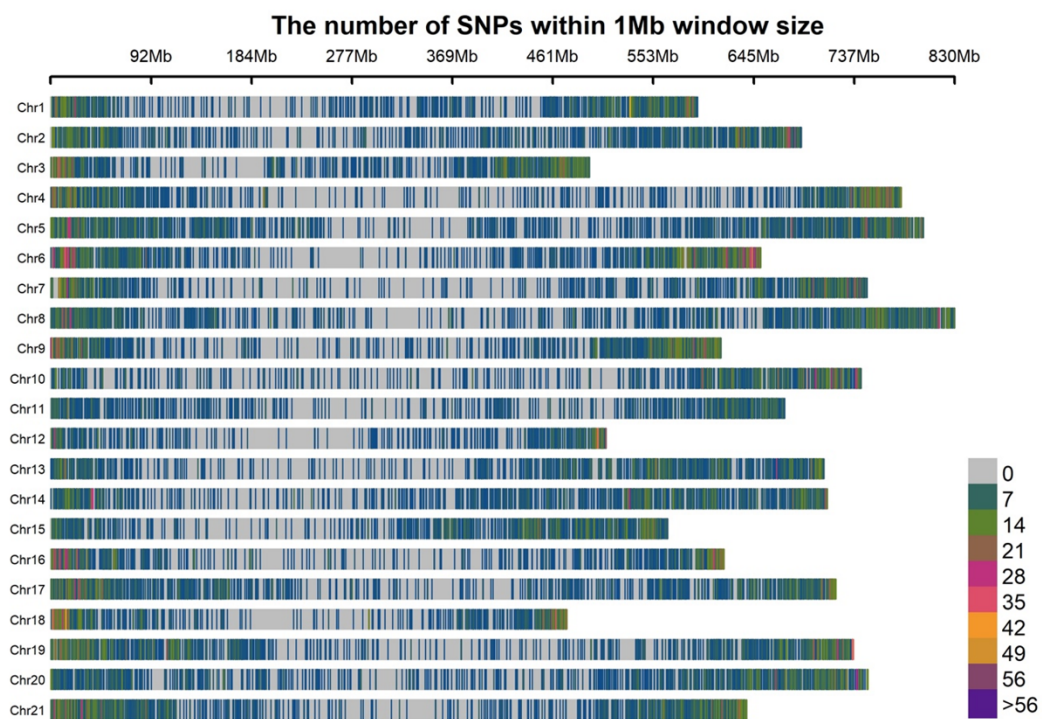
<sup>b</sup>PVE: Phenotypic variance explained



## Figures



**Fig. 1** Frequency distribution of common bunt infected spikes (%) obtained from best linear unbiased predictors combined over two years (2016 and 2017) from 125 synthetic hexaploid wheats.



**Fig. 2** Physical distribution of 35,798 genotyping-by-sequencing derived SNPs within 1-Mb window size on 21 chromosomes of 125 synthetic hexaploid wheats.

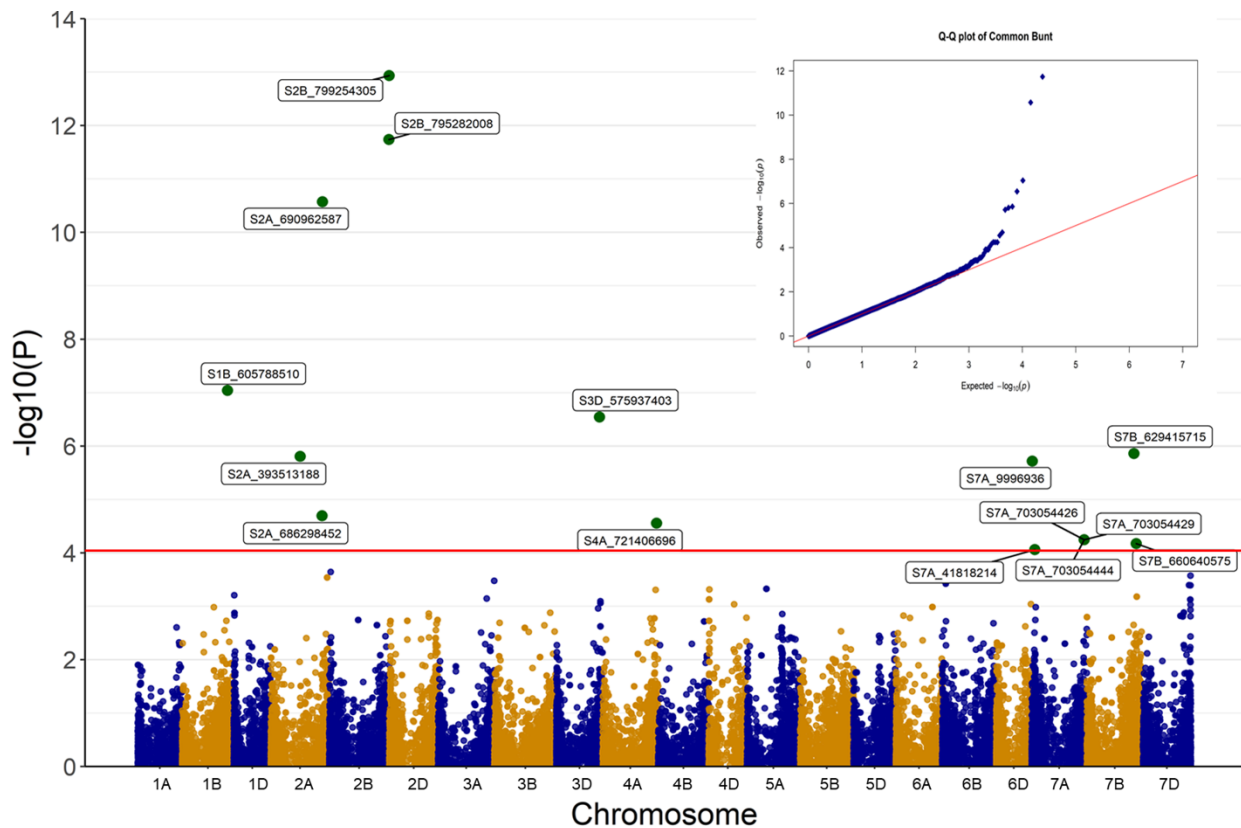


Fig. 3 Manhattan (main panel) and quantile-quantile (Q-Q; top right) plots showing genome-wide association results for common bunt resistance in 125 synthetic hexaploid wheat lines. The Manhattan plot shows the association  $-\log_{10}(p)$  for each genome-wide SNPs (35,798) on y-axis by chromosomal position on x-axis. The green dots in the Manhattan plot shows the significant marker-trait associations above the threshold line (red) with  $P = 8.8E-05$  [ $-\log_{10}P=4.05$ ]. The Q-Q plot shows the deviation of association test statistics (blue dots) from the distribution expected under the null hypothesis (red line).

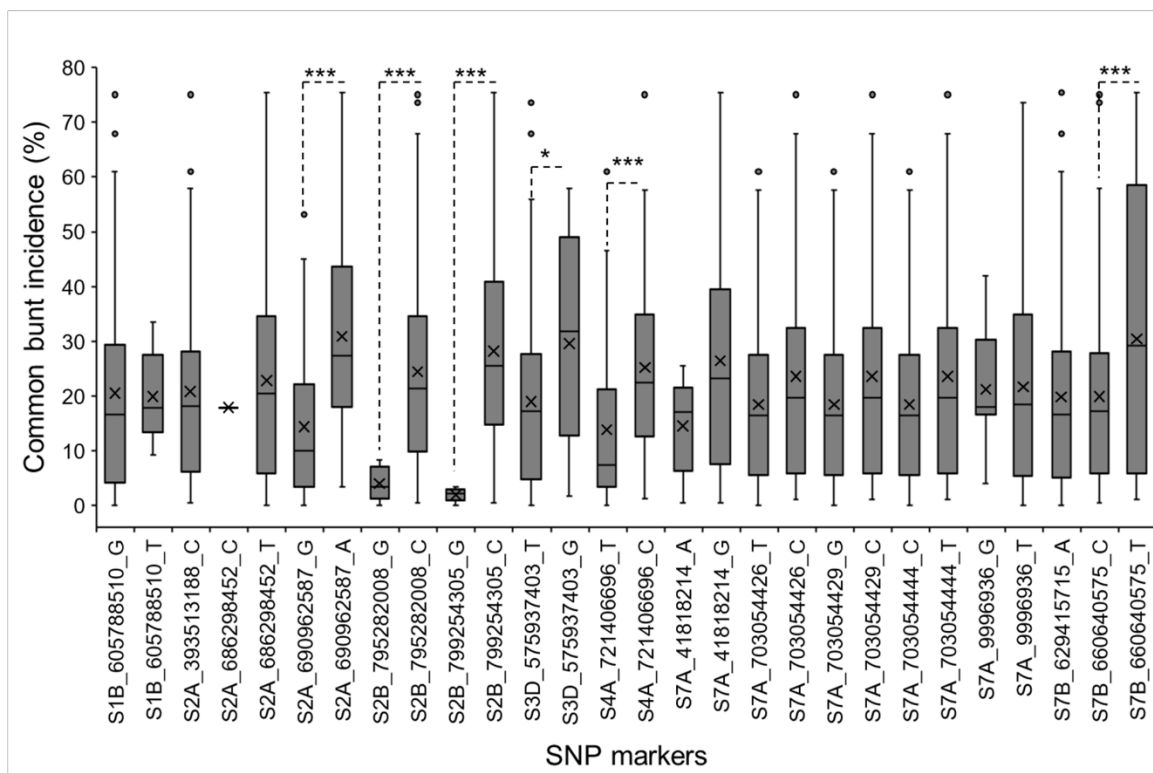


Fig. 4 Distribution of favorable and unfavorable alleles in 125 synthetic hexaploid wheat lines and comparison of alleles of the SNP markers associated with common bunt incidence (%) to determine the significant differences between the mean values of two alleles. The cross symbol and horizontal line inside the box plots are mean and median values of common bunt incidence, respectively. \*; \*\*\* Indicate significance at the 0.05 and 0.001 probability levels, respectively.

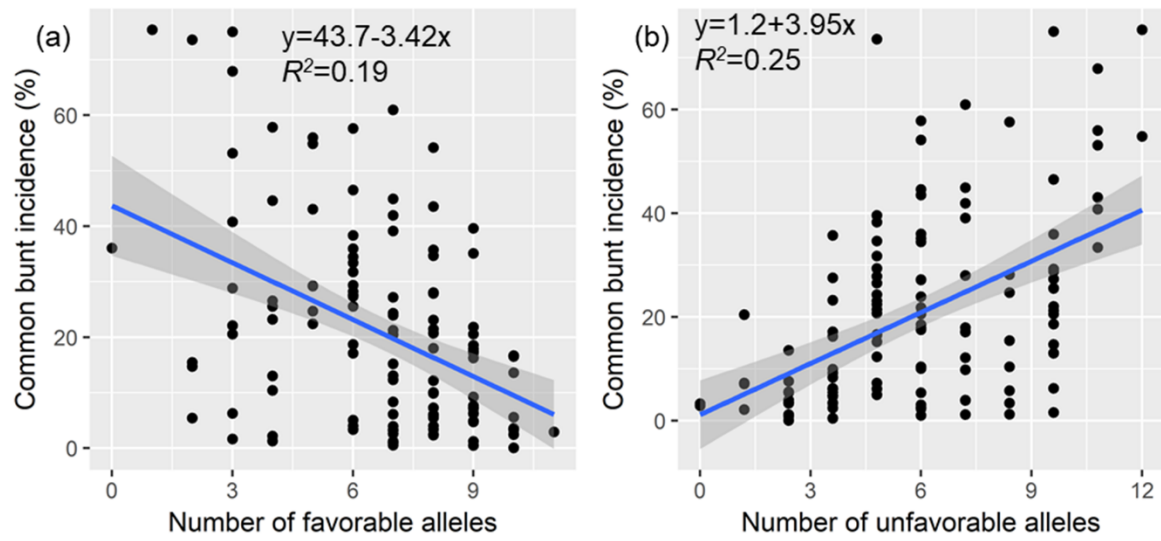


Fig. 5 Scatter plots and regression analysis of common bunt incidence (%) and the number of (a) favorable alleles or (b) unfavorable alleles in 125 synthetic hexaploid wheat lines.

**CHAPTER 5. A GENOME-WIDE ASSOCIATION STUDY REVEALS NOVEL GENOMIC REGIONS ASSOCIATED WITH 10 GRAIN MINERALS IN SYNTHETIC HEXAPLOID WHEAT**

This chapter is published: Bhatta M., Baenziger P.S., Waters B.M., Poudel R., Belamkar V., Poland J., Morgounov A., *International Journal of Molecular Sciences* 19(10):3237 (2018). <https://doi.org/10.3390/ijms19103237>

**Abstract**

Synthetic hexaploid wheat (SHW; *Triticum durum* L. x *Aegilops tauschii* Coss.) is a means of introducing novel genes/genomic regions into bread wheat (*T. aestivum* L.) and a potential genetic resource for improving grain mineral concentrations. We quantified 10 grain minerals (Ca, Cd, Cu, Co, Fe, Li, Mg, Mn, Ni, and Zn) using inductively coupled mass spectrometer in 123 SHWs for genome-wide association study (GWAS). A GWAS with 35,648 SNP markers identified 92 marker-trait associations (MTAs), of which 60 were novel and 40 were within genes, and the genes underlying 20 MTAs had annotations suggesting a potential role in grain mineral concentration. Twenty-four MTAs on the D-genome were novel and showed the potential of *Ae. tauschii* for improving grain mineral concentrations such as Ca, Co, Cu, Li, Mg, Mn, and Ni. Interestingly, the large number of novel MTAs (36) were identified on the AB genome of these SHWs indicated that there is a lot of variation yet to be explored and to be used in the A and B genome along with the D-genome. Regression analysis identified a positive correlation between a cumulative number of favorable alleles in a genotype and grain mineral concentration. Additionally, we identified multi-traits and stable MTAs and recommended 13 top 10% SHWs with a higher concentration of beneficial grain minerals (Cu, Fe, Mg, Mn, Ni, and Zn) and a large number of favorable alleles compared to low ranking genotypes and checks that could be utilized in the breeding program for the genetic

biofortification. This study will further enhance our understanding of the genetic architecture of grain minerals in wheat and related cereals.

Keywords: *Triticum durum*, *Aegilops tauschii*, *Triticum aestivum*, marker-trait associations, genes, bread wheat, genetic biofortification, favorable alleles

## Introduction

The global population is increasing rapidly and is expected to reach 9.8 billion in 2050 [1]. With the increase in global population, the demand for staple crops will continue to increase. Wheat (*Triticum aestivum* L.) is one of the most important staple crops, and it feeds more than one-third of the world population, providing carbohydrates, proteins, vitamins, antioxidants, fibers, and minerals [2]. In 2017/2018, wheat production was estimated at 756.8 million tons [3]. Despite the significant growth in wheat production, a large percentage of the population who rely on wheat as a staple crop suffer from deficiencies in minerals such as calcium (Ca), copper (Cu), iron (Fe), magnesium (Mg), and zinc (Zn) [4–6] because of the of low grain mineral concentrations [7]. Increased concentrations of essential minerals and decreased concentrations of toxic minerals such as cadmium (Cd) in wheat grain will have a significant impact on human health. One sustainable and cost-effective approaches to increasing essential mineral concentrations is through genetic biofortification, which requires identification of cultivars with useful genetic variability for grain minerals and understanding of the physiological and genetic architecture of these minerals in wheat [8].

Grain mineral concentration is dependent on several processes, including mineral absorption from the soil, uptake by the roots, translocation, assimilation, and remobilization to

the seed [9]. The involvement of several processes for the accumulation of minerals in grain makes them complex traits, which are most likely controlled by many genes [8]. Quantitative trait loci (QTL) analysis or genome-wide association study (marker-trait associations; MTAs) approaches are widely used to dissect complex traits. In wheat to date, 13 QTLs and 485 MTAs were identified for Ca [4,8,10], one QTL and 13 MTAs identified for Cd [11,12], 17 QTLs for Cu [8,10,13], 58 QTLs for Fe [5,8,10,13–20], three QTLs for Mg [8,10], 15 QTLs for manganese (Mn) [10,13], and 46 QTLs and 16 MTAs for Zn [5,8,10,13–21]. The identification of QTLs or MTAs for high concentrations of beneficial grain minerals such as Ca, Cu, Cobalt (Co), Fe, Lithium (Li), Mg, Mn, Nickel (Ni), and Zn, and low Cd concentration will assist in genetic biofortification through marker-assisted selection and ultimately assist in ensuring nutritional security.

Improved wheat cultivars contain low concentrations of grain minerals [5] and have narrow genetic variation for grain minerals compared to wheat's wild relatives [22]. Synthetic hexaploid wheat (SHW; *Triticum turgidum* L. x *Aegilops tauschii* Coss.) is being used as a means of introducing novel genes/genetic variation into bread wheat [23,24] and it is a potential source of high grain mineral concentrations [25]. Thus, we selected a panel of 123 synthetic hexaploid wheat genotypes to (i) explore the genetic variation of 10 grain minerals (Ca, Cd, Co, Cu, Fe, Li, Mg, Mn, Ni, and Zn) and grain protein concentration (GPC); (ii) identify marker-trait associations using a genome-wide association study and (iii) candidate genes containing nucleotide variants influencing grain minerals. This report is the first for Cu, Co, Fe, Li, Mg, Mn, and Ni in wheat. Results of this study will facilitate the selection of SHWs for use in wheat improvement programs and in enhancing the nutritive value through the integration of valuable grain mineral favorable alleles from SHWs to meet current and future dietary needs.



## Materials and methods

### Plant materials and experimental design

The detail of the experimental materials and design were described previously [26]. In brief, a diversity panel of 123 SHWs originating from CIMMYT, Mexico and Kyoto University, Japan were used (**APPENDIX I**). Grain samples from each plot were obtained from field trials conducted in 2016 and 2017 growing seasons at the research farm located at the Bahri Dagdas International Agricultural Research Institute in Konya, Turkey (37°51'15.894" N, 32°34'3.936" E; Elevation = 1,021 m). The mean monthly temperature in both growing seasons was similar [26], however, the total rainfall in 2017 growing season (243 mm) was slightly higher than that observed in 2016 growing season (222 mm) [26]. However, rainfall in both growing seasons was below 25-years average (435 mm) suggesting the presence of drought stressed environmental conditions [26]. The soil texture was clayey loam, with a mean pH of 7.7 in 2016 and 8.2 in 2017 growing season (**APPENDIX IX**). Details on soil analysis are provided in APPENDIX IX. The experimental design in the 2016 growing season was an augmented design with replicated checks ('Gerek' and 'Karahan') and modified alpha lattice design with replicated checks (Gerek and Karahan) and two replications in 2017 as described previously [26].

### Grain yield, thousand kernel weight, grain protein concentration, and grain mineral analysis

Grain yield (GY), thousand kernel weight (TKW), and grain protein concentration (GPC) were measured using previously reported protocols [26–28]. Whole grain mineral analysis was performed as described previously [12]. In brief, approximately 2 g of oven dried grains were digested with concentrated nitric acid (Optima, Fisher Chemical, Thermo Fisher Scientific Inc.) and hydrogen peroxide (30% H<sub>2</sub>O<sub>2</sub>, Fisher BioReagents, Thermo Fisher Scientific Inc.). Each

digestion set of 50 samples included a reagent blank and 0.25 g of standard reference flour (Standard reference material 1567a, National Bureau of Standards, MD). Grain mineral concentrations were determined in duplicate by inductively coupled plasma-mass spectrometry (ICP-MS; Agilent 7500cx, Agilent Technologies Inc.) with Ar Carrier and a He collision cell at the University of Nebraska Redox Biology Center Proteomics and Metabolomics Core. Mineral concentrations for Ca, Cd, Co, Cu, Fe, Li, Mg, Mn, Ni, and Zn were averaged over the duplicates and a reagent blank was subtracted. Mineral concentrations expressed as mg Kg<sup>-1</sup> (dry weight basis) were used for further analysis [29].

#### Phenotypic data analysis

Combined over two years and individual year analyses of variance (ANOVA) were computed using a mixed linear model using PROC MIXED in SAS 9.4 [30]. This was performed to estimate the best linear unbiased predictors (BLUPs) and to determine whether significant variations exist among the genotype, year, and their interactions. The details of the mixed linear model used for the analysis were described previously [26]. In brief, for the combined ANOVA, year and check were assumed as fixed effects whereas genotype, genotype x year interaction, replication nested within a year, and incomplete block nested within replications were assumed as random effects. For augmented design in 2016, ANOVA was calculated by assuming check as a fixed effect whereas genotype and incomplete block as random effects. Incomplete blocks nested within replication, checks fitted into new variable (new variable: check was coded as 0 and entry was coded as 1, where genotype was taken as a new variable x entry), and replications were used to correct for spatial variation in the data. For modified alpha ( $\alpha$ ) lattice design in 2017, ANOVA was calculated by assuming check as a fixed effect and genotype, replication,

and incomplete block nested within replication as random effects. Broad-sense heritability was calculated based on entry mean basis using following formula:

$$H^2 = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{gxyr}}{n} + \frac{\sigma^2_e}{nr}} \quad (1)$$

Where,  $\sigma^2_g$ ,  $\sigma^2_e$ , and  $\sigma^2_{gxyr}$  are the variance components for genotype, error, and genotype x year, respectively, whereas n and r are the number of years and replications, respectively.

The phenotypic correlation was computed using PROC CORR in SAS using BLUPs of each trait. To understand the association among grain minerals, GPC, and GY, a factor analysis using principal component (PC) method and varimax rotation was performed on correlation matrix in each year using the factoextra package in R software [31]. Furthermore, Canonical correlation was performed between GY/GPC and mineral concentrations to determine the relationship between GY/GPC and overall mineral concentration.

#### Genotyping and SNP discovery

Genotyping, SNP discovery, and SNP filtering procedures were described previously [23]. Briefly, DNA was extracted from fresh young leaves (approx. 14 days after sowing) using BioSprint 96 Plant Kits (Qiagen, Hombrechtikon, Switzerland). Genotyping was performed with high-density markers using a genotyping-by-sequencing approach [32]. SNP discovery was performed using TASSEL v. 5.2.40 GBS v2 Pipeline [33] with a physical alignment to the Chinese spring genome sequence (RefSeq v1.0) provided by the International Wheat Genome Sequencing Consortium (IWGSC) [34]. The identified SNPs were filtered for minor allele frequency (MAF > 5%) and missing data (<20%) [23,35].

### Population structure and genome-wide association analysis

The population structure analysis was described previously [23]. In brief, the population structure of 123 genotypes was assessed using STRUCTURE v 2.3.4 [36] and unweighted pair group method with arithmetic mean using TASSEL [37].

A GWAS was performed separately for each year [BLUPs from 2016 (BLUP16), and BLUPs from 2017 (BLUP17), and BLUPs combined over years (CBLUP)] to identify MTAs for grain minerals using FarmCPU (Fixed and random model Circulating Probability Unification) with population structure ( $Q_{1-3}$ ) as a fixed effect (covariate) and FarmCPU calculated kinship as a random effect [38] implemented in the MVP R software package (<https://github.com/XiaoleiLiuBio/MVP>). The identified MTAs were tested against a Bonferroni correction at 5% level of significance with a  $P = 1.4026E-06$  ( $-\log_{10}P = 5.85$ ) for multiple testing correction. Regression analysis was performed between the cumulative number of favorable alleles in a genotype and the BLUPs of each trait. Functional annotations of genes were retrieved using the IWGSC RefSeq v1.0 annotations provided for Chinese spring (IWGSC) [34]. The impact of nucleotide variants on predicted genes or proteins was investigated using SnpEff software (<http://snpeff.sourceforge.net/>).

## Results and Discussion

### Phenotypic variation for grain protein concentration and grain minerals

Genotypic variability for GPC and grain minerals were assessed in 123 SHWs across two years (2016 and 2017) in field studies in Turkey. The ANOVA combined over years revealed a significant effect of genotype for all traits whereas significant genotype x year effect was observed for GPC, Ca, Cu, Mg, Mn, and Ni (**Table 1**). Non-significant genotype x year

interaction for Cd, Co, Fe, Li, and Zn indicate the genetic stability of these traits across years. A wide range of genotypic variation for GPC and minerals was observed among the 123 SHWs (**Table 1**). A wide range of genetic variation observed for GY and TKW in these SHWs were described previously [26]. Variation for GPC ranged from 130 to 168 g Kg<sup>-1</sup> with an average of 151 g Kg<sup>-1</sup> in 2016 and from 116 to 169 g Kg<sup>-1</sup> with an average of 138 g Kg<sup>-1</sup> in 2017 (**Table 1**). Similarly, variation for grain Fe concentration combined over two years ranged from 17 to 65 mg Kg<sup>-1</sup> with an average of 39 mg Kg<sup>-1</sup> and for grain Zn concentration ranged from 10 to 39 mg Kg<sup>-1</sup> with an average of 23 mg Kg<sup>-1</sup>. Some of these SHWs had higher grain Co, Cu, Fe, Li, and Mg concentrations; similar grain concentrations of Mn, Ni, and Zn, and lower grain Cd and Ca concentrations than previously reported in Hard Winter Wheat Association Mapping Panel (HWWAMP) consisting of 299 diverse genotypes representing the USA Great Plains [12]. The lower concentration of grain Ca in the SHWs than in bread wheat cultivars has been reported previously [25]. A previous study had reported much higher grain Cd concentration (up to 0.6mg Kg<sup>-1</sup>) in winter wheat [12] than our study. The Cd concentration in the SHWs in this study was <0.1 mg kg<sup>-1</sup> which is below the regulatory toxic level of 0.2 mg Kg<sup>-1</sup>. However, the low Cd concentration in SHWs may be reflective of low Cd concentration in the soil, and unless they are grown in a high Cd site we cannot ascertain whether these lines will provide low-Cd alleles for breeding [12]. Additionally, a previous study on genetic variation for grain Fe, Mn, and Zn concentrations in SHWs reported between 25-30% higher grain mineral concentrations of Fe, Mn, and Zn than bread wheat cultivars and the higher grain mineral concentrations in SHWs were not only due to lower GYs but also due to a higher nutrient uptake efficiency [25]. This result indicated that the SHWs are potential sources of high grain mineral concentrations and could be used for genetic biofortification of wheat.

Broad-sense heritability estimated across the two years was high ( $H^2 > 0.60$ ) for GPC, Cu, Fe, Mg, Mn, and Zn concentrations; moderate ( $> 0.40$  and  $< 0.60$ ) for Ca and Ni concentrations, and low ( $< 0.40$ ) for Cd, Co, and Li concentrations (**Table 1**). Higher broad sense heritability indicated that the trait was largely governed by the genotypic effect. These results showed potential for the improvement of GPC, Cu, Fe, Mg, Mn, and Zn concentrations through phenotypic selection within SHWs. Similar heritability for these traits has been reported in the previous studies [4,5,8,12,20].

#### Principal component analysis and phenotypic correlation

To understand the association among GY, GPC, and 10 mineral concentrations, a factor analysis using principal component (PC) method was performed in each year (**Figure 1**). The first three PCs explained from 74.6 to 75.8% of the total variation in the data in 2016 and 2017, respectively. In 2016, the first PC explained 53.1% of the variation in the data and the variables included were Ca, Cd, Co, Cu, Fe, Mg, Mn, Ni, and Zn; the second PC explained 13.2% of the variation in data and variables included were GY and GPC; and the third PC explained 8.3% of the variation in the data and variable included was Li. Similarly, in 2017, the first PC explained 54.2% of the variation in the data and variables included were Ca, Cd, Cu, Fe, Mg, Mn, and Zn; the second PC explained 13.2% of the total variance and variables included were GY and GPC; and the third PC explained 8.4% of the total variance and the variables included were Co, Li, and Ni. Most of the grain minerals in both years were included in the PC1 with positive loadings, implying that PC1 is a measure of overall mineral accumulations in the grain, which was similar to the conclusions of Guttieri et al. [12]. The second PC showed a negative correlation between GY and GPC. The association observed in the factor analysis was supported by the significant

positive correlations ( $r$ ) among most of the grain minerals and negative correlation of GY and GPC (**Table 2**).

A significant negative correlation between GY and GPC was reported in previous studies [12,27] and the negative correlation was mainly due to the dilution effect. As expected, the present study also identified a significant negative correlation between GY and GPC (**Table 2**). Additionally, GY was positively correlated with TKW ( $r=0.37$ ,  $p<0.0001$  in 2016 and  $r=0.35$ ,  $p<0.0001$  in 2017) similar to previous studies [26,28]. However, GY was not correlated with grain minerals in this study whereas the significant negative correlation of GY with most of the grain minerals was observed after controlling for TKW (**Table 2**). This result indicated that TKW masked the true association of GY with minerals and controlling for the effect of TKW is important. Furthermore, canonical correlation analysis between GY and overall mineral concentration identified negative correlation ( $r=-0.37$  in 2016 and  $r=-0.16$  in 2017) between them. Similarly, several previous studies have identified a negative correlation of GY with grain minerals, including Fe [8,39] and Zn [8,10,12,39], which were reported to be associated with a dilution effect [12]. In the present study, GPC was significantly positively correlated with Ca, Cd, Cu, Fe, Mg, Mn, Ni, and Zn, however, the correlation was not very strong ( $0.51 \geq r \geq 0.19$ ) (**Table 2**). Additionally, canonical correlation analysis between GPC and overall grain minerals identified positive correlation ( $r=0.44$  in 2016 and  $r=0.62$  in 2017) between them. Many studies have shown a significant positive correlation of GPC with Fe and Zn concentrations [12,15,16,21], indicating that these traits might have a similar genetic basis and could be improved simultaneously [7]. Additionally, most of the grain minerals had highly significant positive correlations ( $p < 0.01$ ) among each other. For instance, strong correlation ( $r > 0.70$ ,  $p < 0.0001$ ) between Fe and Zn was observed, and they were also strongly correlated ( $r > 0.70$ ,  $p <$

0.0001) with other minerals such as Cu, Mg, Mn, and Zn. Positive correlations among grain minerals have been reported previously. For instance, many studies have shown the significant correlation between Fe and Zn concentrations in wheat [5,12,14,16,20,39]. However, other studies have shown no correlation between Fe and Zn [7,40], indicating the genotypic and environmental influence on the relationship between these traits.

Cadmium is a toxic heavy metal that causes harm to human health. Reducing the grain Cd concentration is one of the important plant breeding objectives for creating healthier grains along with the enhancement of beneficial grain mineral concentrations [29]. The current study identified the significant positive correlation between grain Cd concentration with other minerals (**Table 2**). The previous study in HWWAMP (in this case using 286 genotypes) also identified a significant positive correlation between grain Cd and Zn concentration, however, the correlation was not very strong ( $r < 0.49$ ) [12]. However, independent genetic regulation of Cd and Zn has been reported [29] which may help explain this weak correlation. The current study identified weak to moderate correlation ( $r < 0.70$ ) of grain Cd concentration with other grain minerals, implying that enhancement of beneficial mineral concentration may be possible without further increasing grain Cd concentration.

#### Selection of top-ranking genotypes

The 13 top 10% SHW lines were selected from two years of combined data that had higher amounts of GPC and beneficial grain mineral concentrations (Cu, Fe, Mg, Mn, Ni, and Zn) compared to checks and lower ranking genotypes and lower Cd concentration compared to lower ranking genotypes and checks (Gerek and Karahan) (APPENDIX X). For instance, the Fe and Zn concentration in top ranking genotypes ranged from 49.5 to 56.0 mg Kg<sup>-1</sup> and 29 to 35



mg Kg<sup>-1</sup>, respectively, whereas Cd concentration ranged from 0.07 to 0.08 mg Kg<sup>-1</sup> (APPENDIX X). This result indicated that these genotypes could be used in the breeding program as a parent with a goal of increasing beneficial grain minerals for addressing the global mineral deficiencies while decreasing toxic compound such as Cd.

#### Population structure and genome-wide association study

Population structure analysis of the 123 SHWs was performed using 35,648 high quality GBS derived SNPs (MAF>0.05 and missing data <20%) that were well distributed across 21 chromosomes (APPENDIX VI). Our previous study on genetic diversity and population structure analysis of 101 SHWs identified a large amount of novel genetic variation that could be utilized in broadening the genetic diversity of bread wheat germplasms [23]. The population structure analysis identified that the 123 SHWs can be divided into three subgroups as described previously (APPENDIX V) [26].

The substantial genetic diversity in these SHWs and our dense SNP markers [26] could be useful in identifying genetic factors underlying the variation for grain minerals using GWAS. A GWAS analysis performed using a multi-locus mixed linear model implemented in FarmCPU algorithm with 35,648 GBS derived SNPs for 10 grain minerals identified a total of 92 MTAs distributed across 20 chromosomes (**Figure 2**) with phenotypic variance explained (PVE) up to 25% (APPENDIX XI). Thirty-five MTAs were detected on the A genome, 32 MTAs on the B genome, and 25 MTAs on the D-genome of SHWs (**Figure 2**).

### *Calcium*

The 15 MTAs for Ca concentration were observed in 14 different genomic regions on chromosomes 1B, 2B, 2D, 3A, 3B, 3D, 6A, 6B, 7A (**Figure 2**) and the PVE by these MTAs ranged from 2.7 to 21.5% (APPENDIX XI), indicating quantitative nature of inheritance for Ca concentration. Earlier studies have reported QTLs/MTAs for Ca on chromosomes 1A [8,10], 2A [10], 2D [10], 5A [4], 2B, 4A, 4B, 5B, 6A, and 7B [8] in wheat, indicating the involvement of these chromosomes in different mapping populations for Ca concentration. However, it is difficult to align our findings with earlier studies because of the employment of different marker systems such as 90K SNP, short sequence repeat (SSR), and diversity arrays technology (DART), the lack of precise location information in previous literature, or the utilization of a different version of the reference wheat genome than the most recent IWGSC RefSeq v1.0 as described previously [26]. However, the associations identified on the same chromosome as the previous study provided confidence in the reliability of these MTAs. The 11 MTAs identified in this study on chromosomes 1B, 3A, 3B, 3D, 6B, and 7A have not been reported and they are potentially novel MTAs controlling grain Ca concentration. Interestingly, no studies have identified a QTL in the D-genome.

### *Cadmium*

The five MTAs for Cd concentration were observed in five different genomic regions on chromosomes 1A, 2A, 2D, 3A, and 6D (**Figure 2**) with PVE ranging from 1.8 to 14.4% (APPENDIX XI). A previous study on QTL analysis in durum wheat identified a major QTL on chromosome 5B [11]. A GWAS was conducted using 286 winter wheat association mapping population and identified 12 MTAs for Cd on chromosome 5A [29]. All the five MTAs

identified in this study are potentially novel MTAs controlling grain Cd concentration. That our study did not find the QTLs identified in the earlier studies may be due to the complexity of the trait and different genotypes used in this study. The identification of novel MTAs in the D-genome (2D and 6D) clearly represent variation coming from the *Ae. tauschii* and show the potential of SHW for its utilization in a marker-assisted breeding program upon validation in an independent genetic background.

#### *Cobalt, lithium, and nickel*

The present study identified three MTAs on chromosomes 3A, 6D, and 7D for Co, 13 MTAs on chromosomes 1B, 1D, 2A, 2D, 3D, 5A, and 6D for Li, and eight MTAs on chromosomes 1A, 2D, 3A, 4D, 5B, and 6A for Ni (**Figure 2**). There is no previous report on QTL or GWAS analysis for Co, Li, and Ni in wheat. Therefore, all the MTAs identified for Co, Li, and Ni are potentially novel MTAs responsible for Co, Li, and Ni concentrations. Interestingly, our study identified several MTAs on the D-genome for Co, Li, and Ni, which showed the utility of SHWs for the improvement of these traits.

#### *Copper*

A total of 13 MTAs for Cu were identified on chromosomes 1B, 2A, 3A, 3B, 4B, 5A, 5B, 5D, 6A, and 6B (**Figure 2**) with PVE ranging from 1.2 to 17.1% (APPENDIX XI). Earlier studies have identified one QTL for Cu concentration on chromosome 5A in diploid wheat (*T. monococum*) [13], 10 QTLs on chromosomes 1A, 2A, 3B, 4A, 4B, 5A, 6A, 6B, 7A, and 7B in tetraploid wheat [8], and six QTLs on chromosomes 2A, 4A, 4D, 5A, 6A, and 7B in hexaploid

wheat [10]. The five MTAs identified in this study on chromosomes 1B, 3A, 5B, and 5D have not been reported and they are potentially novel MTAs controlling grain Cu concentration.

#### *Iron*

A total of three MTAs for Fe concentration were identified on chromosomes 1A and 3A (**Figure 2**) with PVE ranging from 11.2 to 13.2% (APPENDIX XI). Earlier studies have identified 58 QTLs distributed on 16 chromosomes 1A, 1B, 2A, 2B, 2D, 3A, 3B, 3D, 4B, 4D, 5A, 5B, 6A, 6B, 6D, 7A, 7B, and 7D [5,8,10,13–20].

#### *Magnesium*

A total of 13 MTAs for Mg concentration were identified on chromosomes 1B, 1D, 2D, 3A, 3B, 4A, 4B, 4D, 5B, 5D, and 7A (**Figure 2**) with PVE ranging from 1.4 to 14.6% (APPENDIX XI). Earlier studies have identified eight QTLs for Mg concentration on chromosomes 1B, 2A, 3A, 5B, 6A, 6B, 7A, and 7B in tetraploid wheat [8] and three QTLs on chromosomes 4A, 5A, and 6A in hexaploid wheat [10]. The six MTAs identified in this study on chromosomes 1D, 2D, 3B, 4B, 4D, and 5D have not been reported and they are potentially novel MTAs controlling grain Mg concentration.

#### *Manganese*

A total of six MTAs for Mn concentration were identified on chromosomes 2D, 3A, 4B, 5D, and 6B (**Figure 2**) with PVE ranging from 4.4 to 14.3% (APPENDIX XI). Earlier studies have identified one QTL on chromosome 5A in *T. monoccocum* [13], two QTLs for Mn concentration on chromosomes 2B and 7B in tetraploid wheat [8] and four QTLs on chromosomes 1A, 2B, 3B in hexaploid wheat [10]. All the six MTAs identified in this study on

chromosomes 2D, 3A, 4B, 5D, and 6B have not been reported and they are potentially novel MTAs controlling grain Mn concentration.

#### *Zinc*

A total of 13 MTAs for Zn concentration were identified on chromosomes 1A, 2A, 3A, 3B, 4A, 4B, 5A, and 6B (**Figure 2**) with PVE ranging from 1.8 to 14.1% (APPENDIX XI). Earlier studies have identified 46 QTLs on 15 chromosomes 1A, 1B, 1D, 2A, 2B, 3A, 3D, 4A, 4B, 4D, 5A, 6A, 6B, 7A, and 7B for Zn concentration [5,8,10,13–21]. Additionally, previous GWAS for Zn concentration identified 13 MTAs on chromosomes 1B, 3A, and 4B[29]. Three MTAs identified in this study on chromosome 3B have not been reported and they are potentially novel MTAs controlling grain Zn concentration.

#### Relationship between grain mineral concentrations and number of favorable alleles

The number of favorable alleles in a genotype is the cumulative number of alleles from MTAs that increases the concentration of beneficial minerals whereas decreases the Cd concentration. Linear relationship between grain mineral concentration and number of favorable alleles per genotype was observed (**Figure 3**), implying that the addition of every favorable allele in a genotype contributed to increase beneficial grain mineral concentrations whereas decrease grain Cd concentration. The number of favorable alleles within 123 SHWs ranged from 9 to 37 alleles and variance explained ( $R^2$ ) by favorable alleles on grain minerals ranged from 10 to 53% (**Figure 3**). The top-ranking 13 genotypes have high number of favorable alleles ranging from 23 to 27 alleles (APPENDIX X). This result suggested that pyramiding these favorable alleles can enhance the grain mineral concentrations and be used in a breeding program for the genetic biofortification.

### Multi-trait and stable marker-trait associations

The present study identified common regions associated with multiple traits on chromosomes 1B, 2A, 3A, and 5B (**Figure 2 and APPENDIX XI**). For instance, the MTA for Ca and Mg was identified on chromosome 1B at 6,867,825 bp, Cu and Zn on chromosomes 2A at 742,969,119 bp, Cu and Mg on chromosome 5B at 607,870,649 bp, and Mg, Mn, and Zn on chromosome 3A at 534,469,328 bp. The co-locations of MTAs for Ca, Cu, Mg, Mn, and Zn indicated the same genomic region controlling these traits, which was also supported by highly significant strong positive correlations among those minerals (**Table 2**). These results suggested that the relationship among these traits was at molecular level, indicating a common genetic basis for these traits which could be improved simultaneously. The QTLs co-localization for some of the minerals have previously been reported. The co-localization of grain Zn QTLs with grain Fe QTLs in tetraploid wheat [8] and hexaploid wheat [5,16,20,21], and QTLs for Mn co-located with Fe concentration in tetraploid wheat[8] have been observed. Co-localization may have occurred either by pleiotropy of the same gene involved in controlling mineral concentrations of different elements or by the presence of different linked genes in the same regions controlling mineral concentrations of different elements independently. Although Cd was significantly associated with grain minerals, we did not find co-localization of grain Cd MTAs with the MTAs of other grain minerals, indicating that the grain Cd concentration may be governed by the different genetic mechanism as described previously [29].

In this study, we identified a stable MTA for Ca on chromosome 6B (32,333,184 bp), Cu on 5B (607,870,649 bp), for Mg on 1B (867,825 bp), for Mn on 2D (58,740,285 bp), for Ni on 2D (48,611,294 bp), and for Zn on 3A (534,469,328 bp), and five stable MTAs for Li on chromosomes 1B (606,491,241 bp), 2D (572,031,650 bp), 3D (610,567,350 bp), 5A

(135,164,381 bp), and 6D (30,744,756 bp) (APPENDIX XI ). The stable MTAs identified could be used for the genetic improvement of these traits.

#### Gene underlying marker-trait associations

The MTAs that were identified were searched against the IWGSC RefSeq v1.0 annotation to identify genes underlying the various MTAs identified in this study. Identification of underlying genes with annotations matching the trait function would provide further confidence for these MTAs. The 40 MTAs (Ca: 8 MTAs; Cd:1; Co:2; Cu: 4; Fe: 3; Li: 4; Mg: 5; Mn: 3; Ni: 3; and Zn: 7) for 10 grain minerals were found within genes distributed on chromosomes 1A, 1B, 2A, 2D, 3A, 3B, 4A, 4B, 4D, 5A, 6A, 6B, 6D, and 7A (APPENDIX XII). Of these, 28 MTAs were present in 19 genes whose annotations indicate they are associated with grain minerals (**Table 3** and APPENDIX XII). For instance, MTAs for Fe were located in genes related to Fe concentration such as 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein [41,42], ATP synthase gamma chain [43], F-box family protein domain [42], GDSL esterase/lipase [44], Leucine rich receptor-like protein kinase [45,46], Myb transcription factor [45,46], Na-translocating NADH-quinone reductase subunit A [47]P, No apical meristem (NAM) protein [48], protein DETOXIFICATION [46], ROP guanine nucleotide exchange factor 10 for Fe [42], and universal stress protein family [46]. Additional examples are provided in **Table 3**. This result provides further evidence for these MTAs and indicated that these genes could be important for grain minerals in wheat, however, functional characterization studies are needed to validate the function of these genes.

Furthermore, we identified several MTAs for the same or multiple traits located within genes that had the same gene annotations (**Table 3 and APPENDIX XI**). For instance, some of

the MTAs for the same traits such as Mg on chromosomes on 3B and 4D, and Zn on chromosomes 3B and 6B were within genes that were both annotated as F-box family protein domain. Similarly, some of the MTAs for multiple traits such as Ca (1 MTA) on chromosome 6B, Li (1) on 2D, Mg (1) on 4A, and Zn (1) on 3B were within genes annotated as Leucine-rich repeat receptor-like protein kinase (**Table 3**), indicating that these genes may be important for improving multiple traits. Multiple MTAs for different traits within genes having the same gene annotation was also reported in our previous study on drought stress related-traits [26].

## Conclusions

The SHWs under study are a valuable resource for the genetic improvement of wheat because they were reported to have large amounts of novel genetic diversity including the D-genome diversity [23], are resistant to multiple stresses [24,26], and showed a weak correlation of GY with GPC and most of the minerals, indicating improvement of grain minerals and GPC without sacrificing yield could be possible. Further, the strong positive correlations observed among the most grain minerals suggested the simultaneous improvement of grain minerals could be possible. The top ranking 13 genotypes with higher concentrations of useful grain minerals and lower concentration of Cd identified in this study have large number of favorable alleles and these SHWs could be used as a donor parent in a wheat breeding program for genetic biofortification.

A GWAS identified 92 MTAs and of which 60 MTAs were novel (15 MTAs on the A genome, 21 MTAs on the B genome and 24 MTAs on the D genome). The large number of novel MTAs (36) identified in the AB genome of these SHWs indicated that there is a lot variation yet



to be explored and to be used in the A and B genome. Several MTAs identified in this study were with in genes having potential roles in improving grain mineral concentrations based on information for their annotations in literature, which provided further evidence for the reliability and usefulness of the MTAs identified. However, further investigation on identified genomic regions could significantly assist in genetic biofortification program. Interestingly, this study identified several MTAs for grain minerals located in genes on different chromosomes that had the same gene annotation, suggesting that the same gene family may play a major role in affecting different grain mineral concentrations in SHWs. This study identified multi-trait (Ca and Mg; Cu and Mg; Mg and Cu; Mg, Mn, and Zn) MTAs on chromosomes 1B, 2A, 3A, and 5B which suggested a common genetic basis of these traits showing the possibility of simultaneous improvement of these traits. Additionally, we identified several stable MTAs for Ca, Cu, Li, Mg, Mn, Ni, and Zn that could be used for the genetic improvement of grain minerals. In summary, a wide range of useful genetic variation for grain minerals and identification of several stable, co-localized multi-trait, and novel genomic regions especially on the D-genome demonstrate the potential of SHWs in its utilization in wheat breeding program for the genetic biofortification and this study also provided the information toward further understanding of the genetic complexity of grain minerals accumulation in wheat.

### **Abbreviations used**

SHW, Synthetic Hexaploid Wheat; GY, Grain yield; TKW, Thousand kernel weight; GPC, Grain protein content; TKW, Ca, Calcium; Cd, Cadmium; Co, Cobalt; Cu, Copper; Fe, Iron; Li, Lithium; Mg, Magnesium; Mn, Manganese; Ni, Nickel; Zn, Zinc; GWAS, Genome-wide association study; QTL, Quantitative trait loci; MTA, Marker-trait association; GBS,

Genotyping-by sequencing; BLUP, Best linear unbiased predictor; PCA, Principal component analysis; PVE, Phenotypic variance explained

## References

1. World Population Prospects 2017 Available online:  
<https://esa.un.org/unpd/wpp/Download/Standard/Population/> (accessed on Aug 8, 2018).
2. Poudel, R.; Bhatta, M. Review of Nutraceuticals and Functional Properties of Whole Wheat. *Journal of Nutrition & Food Sciences* 2017, 07, doi:10.4172/2155-9600.1000571.
3. World Food Situation Available online: <http://www.fao.org/worldfoodsituation/csdb/en/> (accessed on Aug 8, 2018).
4. Alomari, D. Z.; Eggert, K.; von Wirén, N.; Pillen, K.; Röder, M. S. Genome-wide association study of calcium accumulation in grains of european wheat cultivars. *Frontiers in Plant Science* 2017, 8, doi:10.3389/fpls.2017.01797.
5. Velu, G.; Tutus, Y.; Gomez-Becerra, H. F.; Hao, Y.; Demir, L.; Kara, R.; Crespo-Herrera, L. A.; Orhan, S.; Yazici, A.; Singh, R. P.; Cakmak, I. QTL mapping for grain zinc and iron concentrations and zinc efficiency in a tetraploid and hexaploid wheat mapping populations. *Plant and Soil* 2017, 411, 81–99, doi:10.1007/s11104-016-3025-8.
6. White, P. J.; Broadley, M. R. Biofortification of crops with seven mineral elements often lacking in human diets – iron, zinc, copper, calcium, magnesium, selenium and iodine. *New Phytologist* 2009, 182, 49–84, doi:10.1111/j.1469-8137.2008.02738.x.
7. Welch, R. M.; Graham, R. D. Breeding for micronutrients in staple food crops from a human nutrition perspective. *Journal of Experimental Botany* 2004, 55, 353–364, doi:10.1093/jxb/erh064.

8. Peleg, Z.; Cakmak, I.; Ozturk, L.; Yazici, A.; Jun, Y.; Budak, H.; Korol, A. B.; Fahima, T.; Saranga, Y. Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat × wild emmer wheat RIL population. *Theoretical Applied Genetics* 2009, *119*, 353–369, doi:10.1007/s00122-009-1044-z.
9. Welch, R. M. Importance of seed mineral nutrient reserves in crop growth and development. In *Mineral Nutrition of Crops: Fundamental Mechanisms and Implications*; Rengel, Z., Ed.; CRC Press, 1999; pp. 205–226 ISBN 978-1-56022-880-6.
10. Shi, R.; Tong, Y.; Jing, R.; Zhang, F.; Zou, C. Characterization of quantitative trait loci for grain minerals in hexaploid wheat (*Triticum aestivum* L.). *Journal of Integrative Agriculture* 2013, *12*, 1512–1521, doi:10.1016/S2095-3119(13)60559-6.
11. Knox, R. E.; Pozniak, C. J.; Clarke, F. R.; Clarke, J. M.; Houshmand, S.; Singh, A. K. Chromosomal location of the cadmium uptake gene (*Cdul*) in durum wheat. *Genome* 2009, *52*, 741–747, doi:10.1139/G09-042.
12. Guttieri, M. J.; Baenziger, P. S.; Frels, K.; Carver, B.; Arnall, B.; Waters, B. M. Variation for grain mineral concentration in a diversity panel of current and historical great plains hard winter wheat germplasm. *Crop Science* 2015, *55*, 1035–1052, doi:10.2135/cropsci2014.07.0506.
13. Ozkan, H.; Brandolini, A.; Torun, A.; Altintas, S.; Eker, S.; Kilian, B.; Braun, H. J.; Salamini, F.; Cakmak, I. Natural variation and identification of microelements content in seeds of einkorn wheat (*Triticum Monococcum*). In *Wheat Production in Stressed Environments*; Developments in Plant Breeding; Springer, Dordrecht, 2007; pp. 455–462 ISBN 978-1-4020-5496-9.

14. Tiwari, V. K.; Rawat, N.; Chhuneja, P.; Neelam, K.; Aggarwal, R.; Randhawa, G. S.; Dhaliwal, H. S.; Keller, B.; Singh, K. Mapping of quantitative trait loci for grain iron and zinc concentration in diploid A genome wheat. *Journal of Heredity* 2009, *100*, 771–776, doi:10.1093/jhered/esp030.
15. Genc, Y.; Verbyla, A. P.; Torun, A. A.; Cakmak, I.; Willsmore, K.; Wallwork, H.; McDonald, G. K. Quantitative trait loci analysis of zinc efficiency and grain zinc concentration in wheat using whole genome average interval mapping. *Plant Soil* 2009, *314*, 49, doi:10.1007/s11104-008-9704-3.
16. Krishnappa, G.; Singh, A. M.; Chaudhary, S.; Ahlawat, A. K.; Singh, S. K.; Shukla, R. B.; Jaiswal, J. P.; Singh, G. P.; Solanki, I. S. Molecular mapping of the grain iron and zinc concentration, protein content and thousand kernel weight in wheat (*Triticum aestivum* L.). *PLoS ONE* 2017, *12*, e0174972, doi:10.1371/journal.pone.0174972.
17. Srinivasa, J.; Arun, B.; Mishra, V. K.; Singh, G. P.; Velu, G.; Babu, R.; Vasistha, N. K.; Joshi, A. K. Zinc and iron concentration QTL mapped in a *Triticum spelta* × *T. aestivum* cross. *Theoretical Applied Genetics* 2014, *127*, 1643–1651, doi:10.1007/s00122-014-2327-6.
18. Roshanzamir, H.; Kordenaeej, A.; Bostani, A. Mapping QTLs related to Zn and Fe concentrations in bread wheat (*Triticum aestivum*) grain using microsatellite markers. *Iranian Journal of Genetics and Plant Breeding* 2013, *2*, 10-17.
19. Distelfeld, A.; Cakmak, I.; Peleg, Z.; Ozturk, L.; Yazici, A. M.; Budak, H.; Saranga, Y.; Fahima, T. Multiple QTL-effects of wheat Gpc-B1 locus on grain protein and

- micronutrient concentrations. *Physiologia Plantarum* 2007, 129, 635–643, doi:10.1111/j.1399-3054.2006.00841.x.
20. Crespo-Herrera, L. A.; Velu, G.; Singh, R. P. Quantitative trait loci mapping reveals pleiotropic effect for grain iron and zinc concentrations in wheat. *Annals of Applied Biology* 2016, 169, 27–35, doi:10.1111/aab.12276.
21. Xu, Y.; An, D.; Liu, D.; Zhang, A.; Xu, H.; Li, B. Molecular mapping of QTLs for grain zinc, iron and protein concentration of wheat across two environments. *Field Crops Research* 2012, 138, 57–62, doi:10.1016/j.fcr.2012.09.017.
22. Cakmak, I. Enrichment of cereal grains with zinc: Agronomic or genetic biofortification? *Plant Soil* 2008, 302, 1–17, doi:10.1007/s11104-007-9466-3.
23. Bhatta, M.; Morgounov, A.; Belamkar, V.; Poland, J.; Baenziger, P. S. Unlocking the novel genetic diversity and population structure of synthetic hexaploid wheat. *BMC Genomics* 2018, 19, 591, doi:10.1186/s12864-018-4969-2.
24. Morgounov, A.; Abugalieva, A.; Akan, K.; Akın, B.; Baenziger, S.; Bhatta, M.; Dababat, A. A.; Demir, L.; Dutbayev, Y.; Bouhssini, M. E.; Erginbaş-Orakci, G.; Kishii, M.; Keser, M.; Koç, E.; Kurespek, A.; Mujeeb-Kazi, A.; Yorgancılar, A.; Özdemir, F.; Öztürk, I.; Payne, T.; Qadimaliyeva, G.; Shamanin, V.; Subasi, K.; Suleymanova, G.; Yakişir, E.; Zelenskiy, Y. High-yielding winter synthetic hexaploid wheats resistant to multiple diseases and pests. *Plant Genetic Resources* 2018, 16, 273–278, doi:10.1017/S147926211700017X.

25. Calderini, D. F.; Ortiz-Monasterio, I. Are synthetic hexaploids a means of increasing grain element concentrations in wheat? *Euphytica* 2003, *134*, 169–178, doi:10.1023/B:EUPH.0000003849.10595.ac.
26. Bhatta, M.; Morgounov, A.; Waters, B. M.; Poudel, R.; Baenziger, P. S. Genome-wide association study reveals novel genomic regions for grain yield and yield-related traits in drought-stressed synthetic hexaploid wheat germplasm. *International Journal of Molecular Sciences* 2018, *19*, 3011. doi:10.3390/ijms19103011.
27. Bhatta, M.; Regassa, T.; Rose, D. J.; Baenziger, P. S.; Eskridge, K. M.; Santra, D. K.; Poudel, R. Genotype, environment, seeding rate, and top-dressed nitrogen effects on end-use quality of modern Nebraska winter wheat. *Journal of the Science of Food and Agriculture* 2017, *97*, 5311–5318, doi:10.1002/jsfa.8417.
28. Bhatta, M.; Eskridge, K. M.; Rose, D. J.; Santra, D. K.; Baenziger, P. S.; Regassa, T. Seeding rate, genotype, and topdressed nitrogen effects on yield and agronomic characteristics of winter wheat. *Crop Science* 2017, *57*, 951–963, doi:10.2135/cropsci2016.02.0103.
29. Guttieri, M. J.; Baenziger, P. S.; Frels, K.; Carver, B.; Arnall, B.; Wang, S.; Akhunov, E.; Waters, B. M. Prospects for selecting wheat with increased zinc and decreased cadmium concentration in grain. *Crop Science* 2015, *55*, 1712, doi:10.2135/cropsci2014.08.0559.
30. SAS 9.4 Product Documentation Available online: <https://support.sas.com/documentation/94/> (accessed on Aug 16, 2018).
31. R: The R Project for Statistical Computing Available online: <https://www.r-project.org/> (accessed on Aug 16, 2018).

32. Poland, J. A.; Brown, P. J.; Sorrells, M. E.; Jannink, J. Development of High-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 2012, 7, doi:10.1371/journal.pone.0032253.
33. Glaubitz, J. C.; Casstevens, T. M.; Lu, F.; Harriman, J.; Elshire, R. J.; Sun, Q.; Buckler, E. S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 2014, 9, e90346, doi:10.1371/journal.pone.0090346.
34. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 2018, 361, eaar7191, doi:10.1126/science.aar7191.
35. Belamkar, V.; Guttieri, M. J.; Hussain, W.; Jarquín, D.; El-basyoni, I.; Poland, J.; Lorenz, A. J.; Baenziger, P. S. Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3 (Bethesda)* 2018, 8, 2735–2747, doi:10.1534/g3.118.200415.
36. Pritchard, J. K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* 2000, 155, 945–959.
37. Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y.; Buckler, E. S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 2007, 23, 2633–2635, doi:10.1093/bioinformatics/btm308.
38. Liu, X.; Huang, M.; Fan, B.; Buckler, E. S.; Zhang, Z. Iterative Usage of Fixed and Random Effect models for powerful and efficient genome-wide association studies. *PLOS Genetics* 2016, 12, e1005767, doi:10.1371/journal.pgen.1005767.



39. Morgounov, A.; Gómez-Becerra, H. F.; Abugalieva, A.; Dzhunusova, M.; Yessimbekova, M.; Muminjanov, H.; Zelenskiy, Y.; Ozturk, L.; Cakmak, I. Iron and zinc grain density in common wheat grown in Central Asia. *Euphytica* 2007, *155*, 193–203, doi:10.1007/s10681-006-9321-2.
40. Graham, R.; Senadhira, D.; Beebe, S.; Iglesias, C.; Monasterio, I. Breeding for micronutrient density in edible portions of staple food crops: conventional approaches. *Field Crops Research* 1999, *60*, 57–80, doi:10.1016/S0378-4290(98)00133-6.
41. Moran Lauter, A. N.; Peiffer, G. A.; Yin, T.; Whitham, S. A.; Cook, D.; Shoemaker, R. C.; Graham, M. A. Identification of candidate genes involved in early iron deficiency chlorosis signaling in soybean (*Glycine max*) roots and leaves. *BMC Genomics* 2014, *15*, 702, doi:10.1186/1471-2164-15-702.
42. Li, W.; Lan, P. Genome-wide analysis of overlapping genes regulated by iron deficiency and phosphate starvation reveals new interactions in Arabidopsis roots. *BMC Research Notes* 2015, *8*, doi:10.1186/s13104-015-1524-y.
43. Laganowsky, A.; Gómez, S. M.; Whitelegge, J. P.; Nishio, J. N. Hydroponics on a chip: Analysis of the Fe deficient Arabidopsis thylakoid membrane proteome. *Journal of Proteomics* 2009, *72*, 397–415, doi:10.1016/j.jprot.2009.01.024.
44. Singh, S. P.; Jeet, R.; Kumar, J.; Shukla, V.; Srivastava, R.; Mantri, S. S.; Tuli, R. Comparative transcriptional profiling of two wheat genotypes, with contrasting levels of minerals in grains, shows expression differences during grain filling. *PLoS ONE* 2014, *9*, e111718, doi:10.1371/journal.pone.0111718.

45. Zheng, L.; Huang, F.; Narsai, R.; Wu, J.; Giraud, E.; He, F.; Cheng, L.; Wang, F.; Wu, P.; Whelan, J.; Shou, H. Physiological and transcriptome analysis of iron and phosphorus interaction in rice seedlings. *Plant physiology* 2009, *151*, 262–274, doi:10.1104/pp.109.141051.
46. King, K. E.; Peiffer, G. A.; Reddy, M.; Lauter, N.; Lin, S. F.; Cianzio, S.; Shoemaker, R. C. Mapping of iron and zinc quantitative trait loci in soybean for association to iron deficiency chlorosis resistance. *Journal of Plant Nutrition* 2013, *36*, 2132–2153, doi:10.1080/01904167.2013.766804.
47. Elias, D. A.; Yang, F.; Mottaz, H. M.; Beliaev, A. S.; Lipton, M. S. Enrichment of functional redox reactive proteins and identification by mass spectrometry results in several terminal Fe(III)-reducing candidate proteins in *Shewanella oneidensis* MR-1. *Journal of Microbiological Methods* 2007, *68*, 367–375, doi:10.1016/j.mimet.2006.09.023.
48. Waters, B. M.; Uauy, C.; Dubcovsky, J.; Grusak, M. A. Wheat (*Triticum aestivum*) NAM proteins regulate the translocation of iron, zinc, and nitrogen compounds from vegetative tissues to grain. *Journal of Experimental Botany* 2009, *60*, 4263–4274, doi:10.1093/jxb/erp257.
49. Liang, W.-W.; Huang, J.-H.; Li, C.-P.; Yang, L.-T.; Ye, X.; Lin, D.; Chen, L.-S. MicroRNA-mediated responses to long-term magnesium-deficiency in *Citrus sinensis* roots revealed by Illumina sequencing. *BMC Genomics* 2017, *18*, doi:10.1186/s12864-017-3999-5.

50. Billard, V.; Maillard, A.; Garnica, M.; Cruz, F.; Garcia-Mina, J.-M.; Yvin, J.-C.; Ourry, A.; Etienne, P. Zn deficiency in *Brassica napus* induces Mo and Mn accumulation associated with chloroplast proteins variation without Zn remobilization. *Plant Physiology and Biochemistry* 2015, *86*, 66–71, doi:10.1016/j.plaphy.2014.11.005.
51. Song, G.; Yuan, S.; Wen, X.; Xie, Z.; Lou, L.; Hu, B.; Cai, Q.; Xu, B. Transcriptome analysis of Cd-treated switchgrass root revealed novel transcripts and the importance of HSF/HSP network in switchgrass Cd tolerance. *Plant Cell Reports* 2018, doi:10.1007/s00299-018-2318-1.
52. Van De Mortel, J. E.; Schat, H.; Moerland, P. D.; Van Themaat, E. V. L.; Van Der Ent, S.; Blankestijn, H.; Ghandilyan, A.; Tsiatsiani, S.; Aarts, M. G. M. Expression differences for genes involved in lignin, glutathione and sulphate metabolism in response to cadmium in *Arabidopsis thaliana* and the related Zn/Cd-hyperaccumulator *Thlaspi caerulescens*. *Plant, Cell & Environment* 2008, *31*, 301–324, doi:10.1111/j.1365-3040.2007.01764.x.
53. Wang, G.; Leonard, J. M.; von Zitzewitz, J.; James Peterson, C.; Ross, A. S.; Riera-Lizarazu, O. Marker–trait association analysis of kernel hardness and related agronomic traits in a core collection of wheat lines. *Molecular Breeding* 2014, doi:10.1007/s11032-014-0028-0.
54. Waters, B. M.; Grusak, M. A. Quantitative trait locus mapping for seed mineral concentrations in two *Arabidopsis thaliana* recombinant inbred populations. *New Phytologist* 2008, *179*, 1033–1047, doi:10.1111/j.1469-8137.2008.02544.x.

55. Chandran, D.; Sharopova, N.; Ivashuta, S.; Gantt, J. S.; VandenBosch, K. A.; Samac, D. A. Transcriptome profiling identified novel genes associated with aluminum toxicity, resistance and tolerance in *Medicago truncatula*. *Planta* 2008, 228, 151–166, doi:10.1007/s00425-008-0726-0.

## TABLES

Table 1. Analysis of variance and phenotype variation for 10 grain minerals, grain protein content and grain yield with minimum (min), maximum (max), fold change (max/min), mean, and broad sense heritability ( $H^2$ ) values of 123 synthetic hexaploid wheats grown in 2016 and 2017 growing seasons in Konya, Turkey.

Trait	2016				2017				Trials combined			$H^2$
	Min	Max	Fold	Mean	Min	Max	Fold	Mean	Year (Yr)	Genotype (G)	G x Yr	
Ca (mg Kg <sup>-1</sup> )	47.5	167.2	3.5	103.1	21.6	84.5	3.9	44.3	*	**	***	0.41
Cd (mg Kg <sup>-1</sup> )	0.03	0.10	3.44	0.07	0.02	0.13	7.68	0.07	NS	*	NS	0.28
Co (mg Kg <sup>-1</sup> )	0.01	0.06	6.53	0.03	0.01	0.04	6.86	0.02	***	*	NS	0.33
Cu (mg Kg <sup>-1</sup> )	2.8	11.4	4.1	7.5	2.9	8.9	3.1	5.7	NS	***	***	0.63
Fe (mg Kg <sup>-1</sup> )	17.7	61.8	3.5	40.2	15.4	67.7	4.4	38.5	NS	***	NS	0.78
Li (mg Kg <sup>-1</sup> )	0.04	0.23	6.43	0.09	0.13	1.07	8.43	0.52	***	*	NS	0.35
Mg (mg Kg <sup>-1</sup> )	617	2097	3	1391	659	2131	3	1458	NS	***	**	0.62
Mn (mg Kg <sup>-1</sup> )	20.3	66.2	3.3	41.2	21.5	69.8	3.2	44.9	NS	***	**	0.67
Ni (mg Kg <sup>-1</sup> )	0.21	2.22	10.81	0.91	0.13	1.16	8.75	0.48	NS	***	***	0.52
Zn (mg Kg <sup>-1</sup> )	8.8	38.1	4.3	23.1	11.1	39.6	3.6	23	NS	***	NS	0.65
Grain protein content (g Kg <sup>-1</sup> )	129.8	167.6	1.3	151.2	116.4	168.9	1.5	137.8	***	***	**	0.68
Grain yield (g m <sup>-2</sup> )	54.3	530	9.8	259	194.7	479.5	2.5	290.1	NS	*	*	0.44

\*, \*\*, and \*\*\* Significant at the 0.05, 0.01, and 0.001 probability level, respectively; NS: Non-significant at the 0.05 probability level.

152 Table 2. Pearson's correlation coefficients of 10 grain minerals, grain protein content (GPC), grain yield (GY), and GY controlling for thousand kernel weight ( $GY_{pTKW}$ ) in 123 synthetic hexaploid wheat grown in 2016 (upper triangle) and 2017 (lower triangle) growing seasons in Konya, Turkey.

Trait	Ca	Cd	Co	Cu	Fe	Li	Mg	Mn	Ni	Zn	GPC	GY	$GY_{pTKW}$
CA	1	0.63***	0.42***	0.64***	0.58***	0.31***	0.80***	0.79***	0.44***	0.60***	0.36***	-0.01	-0.11
CD	0.64***	1	0.32***	0.68***	0.61***	0.38***	0.65***	0.67***	0.58***	0.63***	0.22*	-0.03	-0.40***
CO	0.37***	0.34***	1	0.49***	0.63***	0.19*	0.49***	0.49***	0.54***	0.53***	-0.05	0.13	0.01
CU	0.82***	0.67***	0.42***	1	0.79***	0.27**	0.84***	0.81***	0.43***	0.89***	0.22*	0.04	-0.18*
FE	0.80***	0.67***	0.46***	0.89***	1	0.24**	0.79***	0.82***	0.48***	0.84***	0.19*	0.08	-0.13
LI	0.41***	0.33***	0.43***	0.30**	0.30**	1	0.38***	0.26**	0.29**	0.17	0.09	-0.13	-0.14
MG	0.87***	0.66***	0.47***	0.90***	0.89***	0.38***	1	0.88***	0.50***	0.83***	0.20*	-0.01	-0.19*
MN	0.77***	0.64***	0.46***	0.87***	0.87***	0.26**	0.91***	1	0.50***	0.83***	0.19*	-0.08	-0.25**
NI	0.32***	0.27**	0.37***	0.42***	0.38***	0.30***	0.46***	0.34***	1	0.38***	0.21*	-0.02	-0.28**
ZN	0.75***	0.65***	0.32***	0.85***	0.86***	0.14	0.87***	0.85***	0.31***	1	0.23*	0.07	-0.19*
GPC	0.31***	0.22*	0.12	0.43***	0.48***	-0.08	0.40***	0.47***	0.19*	0.51***	1	-0.37***	-0.36***
GY	0.03	0.04	0.02	-0.01	-0.07	0.13	0.05	0.02	-0.03	-0.07	-0.36***	1	-
$GY_{pTKW}$	-0.11	-0.10	0.07	-0.19*	0.22*	0.06	-0.14	-0.15	-0.13	-0.24**	-0.44***	-	1

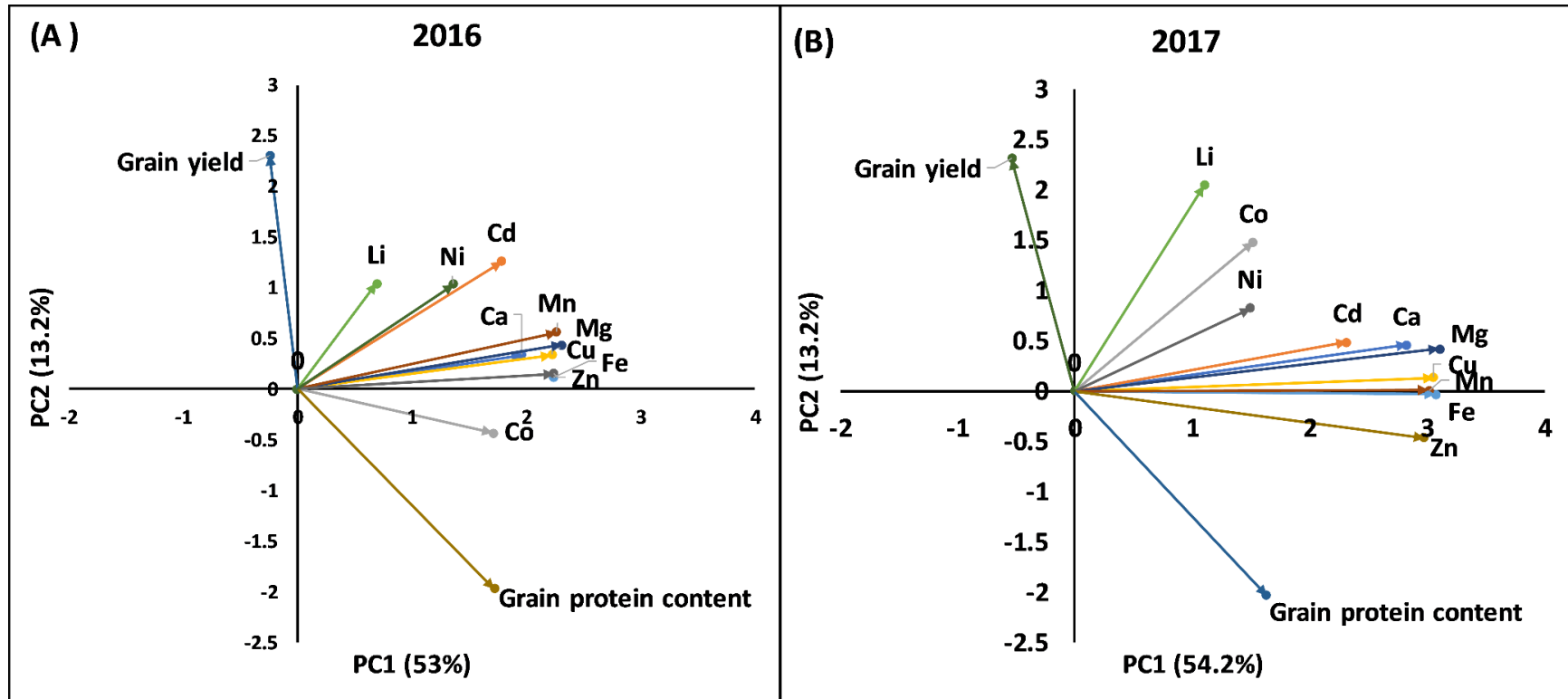
\*, \*\*, and \*\*\* Significant at the 0.05, 0.01, and 0.001 probability level, respectively.

Table 3. Potential candidate genes containing/flanking marker-trait associations for improving grain minerals in SHWs.

Gene annotation (GeneID)	Trait in our study a	Chromosome	PVE (%) <sup>b</sup>	Traits influenced based on the annotations	References for the association of annotations with traits
2-oxoglutarate (2OG) and Fe (II)-dependent oxygenase superfamily protein (TraesCS2A01G519900-TraesCS2A01G520000)	Cu (1)	2A	5.3	Fe, Mg	[42,49]
AP2-like ethylene-responsive transcription factor (TraesCS2A01G514200)	Cu (1)	2A	3.1	As	
ATP synthase gamma chain (TraesCS6B01G117700)	Ca (1)	6B	19.9	Fe, Zn	[50]
Chaperone protein dnaJ (TraesCS4B01G187600)	Zn (1)	4B	14.1	Cd	[41]
F-box family protein domain (TraesCS6D01G360300, TraesCS3B01G479800, TraesCS3B01G11900, TraesCS6B01G268400, TraesCS6D01G064500, TraesCS2D01G106500, TraesCS4D01G333100)	Co (1), Li (1), Mg (2), Mn (1), Zn (2)	2D, 3B, 4D, 6B, 6D	1.8-25.2	Fe	[43]
GDSL esterase/lipase (TraesCS5A01G096300)	Li (1)	5A	4.4	Fe, Zn, Mn	[51]
Kinase family protein (TraesCS1B01G375400)	Li (1)	1B	13.5	Cd, Zn	[44,45]
Leucine rich receptor-like protein kinase (TraesCS2D01G466400, TraesCS6B01G384300-TraesCS6B01G384400, TraesCS4A01G490700, TraesCS3B01G192500)	Ca (1), Li (1), Mg, Zn	2D, 3B, 4A, 6B	1.8-12.6	Fe	[44,45,51]
MYB transcription factor (TraesCS6B01G053100)	Ca (1)	6B	9.9	Cd, Fe, Zn	[46]
Na-translocating NADH-quinone reductase subunit A (TraesCS1A01G432900)	Fe (1)	1A	11.2	Fe	[47]
No apical meristem (NAM) protein (TraesCS7A01G068200)	Ca (1)	7A	11.8	Fe, Zn, N	[51]
Peroxidase (TraesCS6A01G081700)	Ca (1)	6A	9	Cd	[52]
Phosphate translocator (TraesCS3B01G192400)	Zn (1)	3B	1.8	P	[53]
Potassium transporter (TraesCS2D01G106600)	Mn (1)	2D	8.7	K	[54]
Protein COBRA, putative (TraesCS4B01G187300)	Mn (1)	4B	13.4	Al	[45]
Protein DETOXIFICATION (TraesCS3A01G300400)	Mg (1)	3A	14.6	Fe	[55]
Protein ROOT HAIR DEFECTIVE 3 homolog (TraesCS1A01G003300-TraesCS1A01G003400)	Zn (1)	1A	3	Cd	[41]
ROP guanine nucleotide exchange factor 10 (TraesCS4D01G333000)	Mg (1)	4D	7.9	Fe	[45]
Universal stress protein family (TraesCS3B01G418000)	Ca (1)	3B	2.9	Fe, Zn	[46]

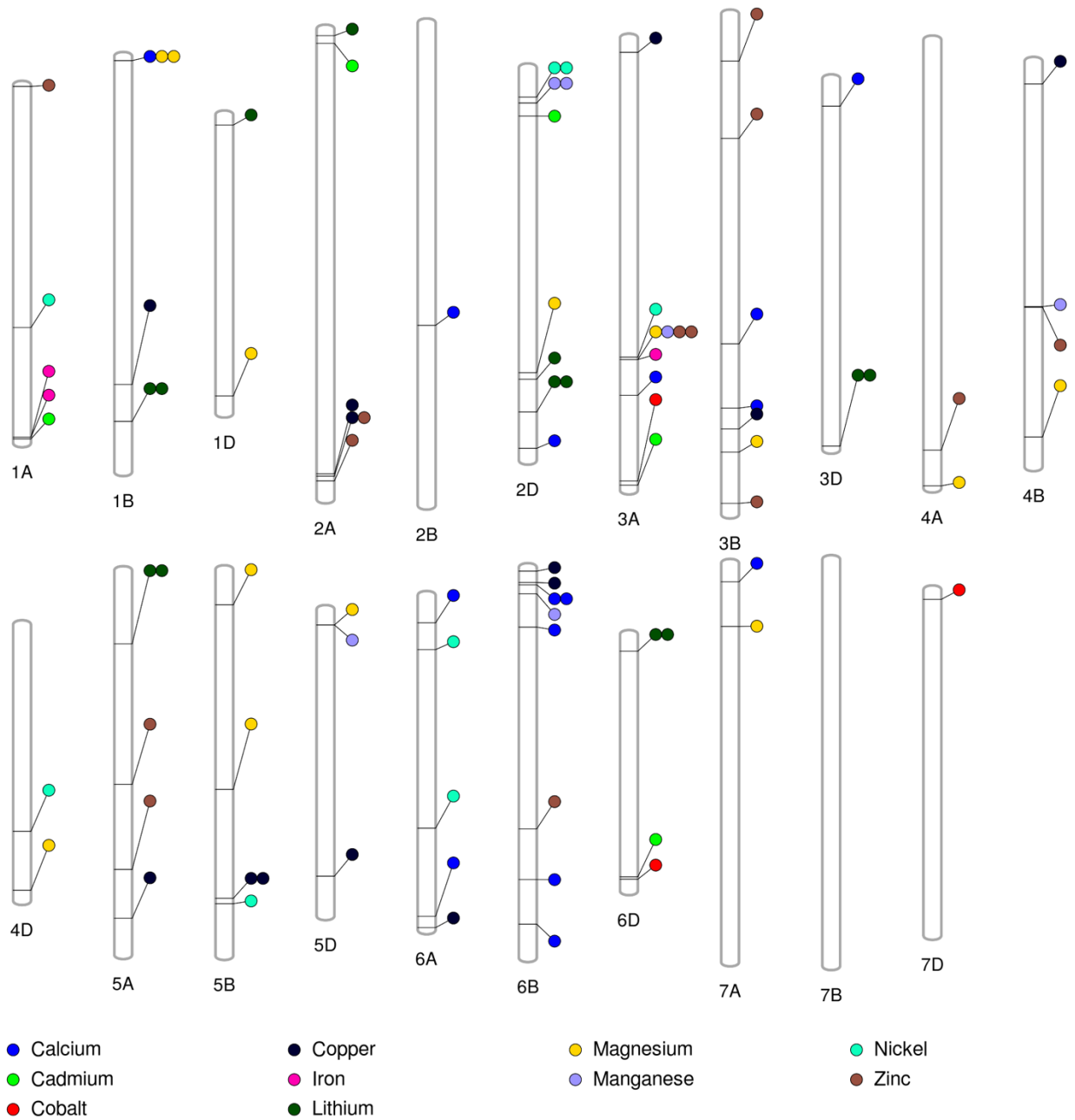
<sup>a</sup> The count of marker-trait associations (in the parenthesis) for either single or multiple traits located within genes that have the same gene annotation; <sup>b</sup> PVE, phenotypic variance explained by the MTA.

## FIGURES

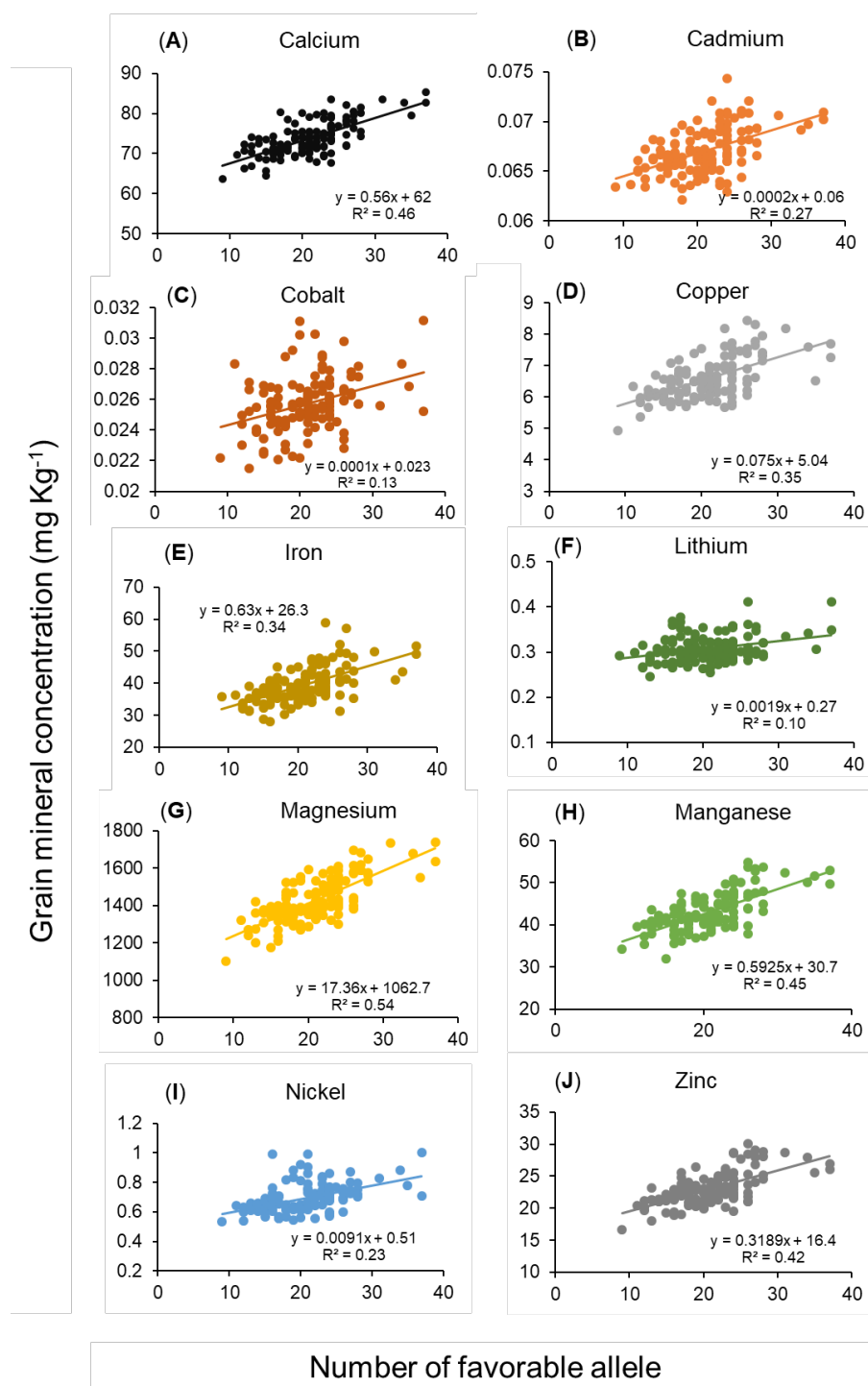


**Figure 1.** Factor analysis using principal component method based on correlation matrix on grain yield, grain protein concentration, and 10 grain mineral concentrations in 123 synthetic hexaploid wheat lines grown in 2016 (A) and 2017 (B) in Konya, Turkey.





**Figure 2.** Significant marker-trait associations identified on each chromosome for 10 grain minerals from a genome-wide association study using 35,648 SNPs in 123 synthetic hexaploid wheat grown in 2016 and 2017 in Konya, Turkey.



**Figure 3.** Regression analysis between total number of favorable alleles per genotype and best linear unbiased predictor values of grain minerals concentrations obtained from two years (2016 and 2017) experiments conducted in Konya, Turkey. Using the MTAs, the number of favorable alleles is defined as the total number of alleles present in a genotype that increases the grain concentration of beneficial minerals such as calcium (A), cobalt (C), copper (D), Iron (E), Lithium (F), Magnesium (G), Manganese (H), Nickel (I), and Zinc (J), whereas decreases grain cadmium (B) concentration.

## APPENDIX

## APPENDIX I. Details of 139 synthetic hexaploid wheats used in this study.

Entry#	Complete Pedigree Information	CID	SELHX	OC	Accession#	SegregationNote
1	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-10E-0E-1YA-0YM	MX-TCI	161641	
2	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-16E-0E-1YA-0YM	MX-TCI	161643	
3	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-18E-0E-3YA-0YM	MX-TCI	161646	
4	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-18E-0E-2YA-0YM	MX-TCI	161645	
5	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-2E-0E-2YA-0YM	MX-TCI	161650	
6	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161649	
7	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-6E-0E-1YA-0YM	MX-TCI	161651	
8	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-10E-0E-1YA-0YM	MX-TCI	161653	8 YELLOW SPIKE
8	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-10E-0E-1YA-0YM	MX-TCI	161653	8 BLACK SPIKE
9	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-11E-0E-1YA-0YM	MX-TCI	161654	
10	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-14E-0E-1YA-0YM	MX-TCI	161656	
11	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-16E-0E-1YA-0YM	MX-TCI	161658	
12	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-1E-0E-1YA-0YM	MX-TCI	161660	
13	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161663	
14	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-3E-0E-1YA-0YM	MX-TCI	161664	
15	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-5E-0E-1YA-0YM	MX-TCI	161666	
16	AISBERG/AE.SQUARROSA(511)	CAWW04GH00005S	-0GH-0SE-030E-8E-0E-1YA-0YM	MX-TCI	161668	

17	LEUC 84693/AE.SQUARROSA(409)	CAWW04GH00010S	-0GH-0SE-030E-10E-0E-3YA-0YM	MX-TCI	161669	17 BROWN SPIKE
17	LEUC 84693/AE.SQUARROSA(409)	CAWW04GH00010S	-0GH-0SE-030E-10E-0E-3YA-0YM	MX-TCI	161669	17 BLACK SPIKE
18	LEUC 84693/AE.SQUARROSA(409)	CAWW04GH00010S	-0GH-0SE-030E-11E-0E-1YA-0YM	MX-TCI	161670	
19	LEUC 84693/AE.SQUARROSA(409)	CAWW04GH00010S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161673	
20	LEUC 84693/AE.SQUARROSA(409)	CAWW04GH00010S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161677	
21	LEUC 84693/AE.SQUARROSA(1026)	CAWW04GH00012S	-0GH-0SE-030E-10E-0E-1YA-0YM	MX-TCI	161679	
22	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-12E-0E-1YA-0YM	MX-TCI	161681	22 BLACK SPIKE
22	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-12E-0E-1YA-0YM	MX-TCI	161681	22 BROWN SPIKE
23	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161687	
24	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-15E-0E-3YA-0YM	MX-TCI	161689	
25	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161694	
26	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-8E-0E-1YA-0YM	MX-TCI	161696	26 BROWN SPIKE
26	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-8E-0E-1YA-0YM	MX-TCI	161696	26 YELLOW SPIKE
27	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-11E-0E-1YA-0YM	MX-TCI	161698	
28	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161700	
29	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-16E-0E-1YA-0YM	MX-TCI	161703	
30	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-18E-0E-1YA-0YM	MX-TCI	161707	
31	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-20E-0E-1YA-0YM	MX-TCI	161709	
32	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-6E-0E-1YA-0YM	MX-TCI	161710	
33	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-8E-0E-1YA-0YM	MX-TCI	161712	
34	UKR-OD 952.92/AE.SQUARROSA(1031)	CAWW04GH00061S	-0GH-0SE-030E-9E-0E-1YA-0YM	MX-TCI	161717	
35	UKR-OD 1530.94/AE.SQUARROSA(310)	CAWW04GH00068S	-0GH-0SE-030E-12E-0E-1YA-0YM	MX-TCI	161719	

37	UKR-OD 1530.94/AE.SQUARROSA(310)	CAWW04GH00068S	-0GH-0SE-030E-6E-0E-1YA-0YM	MX-TCI	161724	
38	UKR-OD 1530.94/AE.SQUARROSA(392)	CAWW04GH00071S	-0GH-0SE-030E-10E-0E-1YA-0YM	MX-TCI	161726	
39	UKR-OD 1530.94/AE.SQUARROSA(392)	CAWW04GH00071S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161730	
40	UKR-OD 1530.94/AE.SQUARROSA(392)	CAWW04GH00071S	-0GH-0SE-030E-23E-0E-1YA-0YM	MX-TCI	161732	
41	UKR-OD 1530.94/AE.SQUARROSA(392)	CAWW04GH00071S	-0GH-0SE-030E-6E-0E-1YA-0YM	MX-TCI	161734	
42	UKR-OD 1530.94/AE.SQUARROSA(392)	CAWW04GH00071S	-0GH-0SE-030E-7E-0E-2YA-0YM	MX-TCI	161736	
43	UKR-OD 1530.94/AE.SQUARROSA(392)	CAWW04GH00071S	-0GH-0SE-030E-7E-0E-1YA-0YM	MX-TCI	161735	
44	UKR-OD 1530.94/AE.SQUARROSA(458)	CAWW04GH00074S	-0GH-0SE-030E-12E-0E-1YA-0YM	MX-TCI	161738	
45	UKR-OD 1530.94/AE.SQUARROSA(458)	CAWW04GH00074S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161740	
46	UKR-OD 1530.94/AE.SQUARROSA(458)	CAWW04GH00074S	-0GH-0SE-030E-4E-0E-1YA-0YM	MX-TCI	161742	
47	UKR-OD 1530.94/AE.SQUARROSA(458)	CAWW04GH00074S	-0GH-0SE-030E-5E-0E-2YA-0YM	MX-TCI	161746	
48	UKR-OD 1530.94/AE.SQUARROSA(458)	CAWW04GH00074S	-0GH-0SE-030E-5E-0E-1YA-0YM	MX-TCI	161745	
49	UKR-OD 1530.94/AE.SQUARROSA(458)	CAWW04GH00074S	-0GH-0SE-030E-9E-0E-1YA-0YM	MX-TCI	161751	
50	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-11E-0E-1YA-0YM	MX-TCI	161752	
51	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-14E-0E-1YA-0YM	MX-TCI	161757	
52	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-18E-0E-1YA-0YM	MX-TCI	161760	52 BLACK SPIKE
52	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-18E-0E-1YA-0YM	MX-TCI	161760	52 BROWN SPIKE
53	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-1E-0E-1YA-0YM	MX-TCI	161765	
54	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-3E-0E-1YA-0YM	MX-TCI		54 YELLOW AWNLESS
54	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-3E-0E-1YA-0YM	MX-TCI		54 BLACK AWNLESS

54	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-3E-0E-1YA-0YM	MX-TCI		54 YELLOW AWNED
54	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-3E-0E-1YA-0YM	MX-TCI	161772	54 BLACK AWNED
55	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-4E-0E-1YA-0YM	MX-TCI	161778	
56	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-11E-0E-1YA-0YM	MX-TCI	161780	
57	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-12E-0E-1YA-0YM	MX-TCI	161782	
58	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-15E-0E-2YA-0YM	MX-TCI	161789	
59	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161788	59 BLACK SPIKE
59	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161788	59 YELLOW SPIKE
60	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-16E-0E-1YA-0YM	MX-TCI	161796	60 BLACK SPIKE
60	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-16E-0E-1YA-0YM	MX-TCI	161796	60 BROWN SPIKE
61	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-17E-0E-3YA-0YM	MX-TCI	161803	
62	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-17E-0E-2YA-0YM	MX-TCI	161802	
63	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-17E-0E-1YA-0YM	MX-TCI	161801	
64	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-18E-0E-1YA-0YM	MX-TCI	161807	
65	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-18E-0E-2YA-0YM	MX-TCI	161808	
66	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161810	
67	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-3E-0E-2YA-0YM	MX-TCI	161813	
68	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-3E-0E-1YA-0YM	MX-TCI	161812	
69	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-5E-0E-2YA-0YM	MX-TCI	161817	
70	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-5E-0E-1YA-0YM	MX-TCI	161816	
71	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-5E-0E-3YA-0YM	MX-TCI	161818	71 BLACK SPIKE
71	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-5E-0E-3YA-0YM	MX-TCI	161818	71 BROWN SPIKE

72	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-7E-0E-2YA-0YM	MX-TCI	161822	
73	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-7E-0E-1YA-0YM	MX-TCI	161821	
74	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-9E-0E-2YA-0YM	MX-TCI	161826	
75	UKR-OD 1530.94/AE.SQUARROSA(1027)	CAWW04GH00078S	-0GH-0SE-030E-9E-0E-1YA-0YM	MX-TCI	161825	
76	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-10E-0E-1YA-0YM	MX-TCI	161827	
77	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-10E-0E-2YA-0YM	MX-TCI	161828	
78	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-13E-0E-1YA-0YM	MX-TCI	161830	
79	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-13E-0E-2YA-0YM	MX-TCI	161831	
80	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-14E-0E-1YA-0YM	MX-TCI	161836	
81	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-16E-0E-1YA-0YM	MX-TCI	161838	
82	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-16E-0E-2YA-0YM	MX-TCI	161839	
83	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-17E-0E-4YA-0YM	MX-TCI	161846	
84	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-17E-0E-1YA-0YM	MX-TCI	161843	
85	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-17E-0E-2YA-0YM	MX-TCI	161844	
86	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-18E-0E-1YA-0YM	MX-TCI	161852	
87	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-2E-0E-2YA-0YM	MX-TCI	161855	
88	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161854	
89	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-4E-0E-2YA-0YM	MX-TCI	161861	
90	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-4E-0E-1YA-0YM	MX-TCI	161860	
91	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-4E-0E-3YA-0YM	MX-TCI	161862	
92	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-9E-0E-1YA-0YM	MX-TCI	161864	
93	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-11E-0E-3YA-0YM	MX-TCI	161868	

94	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-11E-0E-1YA-0YM	MX-TCI	161866	
95	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-11E-0E-2YA-0YM	MX-TCI	161867	
96	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-13E-0E-1YA-0YM	MX-TCI	161874	
97	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-15E-0E-1YA-0YM	MX-TCI	161876	
98	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-17E-0E-2YA-0YM	MX-TCI	161879	98_AWNLESS
98	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-17E-0E-2YA-0YM	MX-TCI	161879	98_AWNED
99	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-17E-0E-1YA-0YM	MX-TCI	161878	
100	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-17E-0E-3YA-0YM	MX-TCI	161880	
101	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-2E-0E-1YA-0YM	MX-TCI	161883	
102	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-4E-0E-1YA-0YM	MX-TCI	161884	
103	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-6E-0E-1YA-0YM	MX-TCI	161885	103_YELLOW SPIKE
103	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-6E-0E-1YA-0YM	MX-TCI	161885	103_BLACK SPIKE
104	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-7E-0E-1YA-0YM	MX-TCI	161889	
105	PANDUR/AE.SQUARROSA(409)	CAWW04GH00081S	-0GH-0SE-030E-7E-0E-2YA-0YM	MX-TCI	161890	
106	LANGDON/IG 48042			JAP		
107	LANGDON/IG 126387			JAP		
108	LANGDON/IG 131606			JAP		
109	LANGDON/KU-2074			JAP		
110	LANGDON/KU-2075			JAP		
111	LANGDON/KU-2088			JAP		
112	LANGDON/KU-2092			JAP		
113	LANGDON/KU-2096			JAP		
114	LANGDON/KU-2097			JAP		
115	LANGDON/KU-2100			JAP		
116	LANGDON/KU-2105			JAP		



117	LANGDON/KU-2079			JAP		
118	LANGDON/KU-20-9			JAP		
119	LANGDON/KU-2093			JAP		
120	LANGDON/PI 508262			JAP		
121	AISBERG/AE.SQUARROSA(369)	CAWW04GH00003S	-0GH-0SE-030E-18E-0E-1YA-0YM	MX-TCI	161644	
122	LEUC 84693/AE.SQUARROSA(409)	CAWW04GH00010S	-0GH-0SE-030E-11E-0E-2YA-0YM	MX-TCI	161671	
123	UKR-OD 1530.94/AE.SQUARROSA(629)	CAWW04GH00076S	-0GH-0SE-030E-11E-0E-2YA-0YM	MX-TCI	161753	
124	AISBERG/AE.SQUARROSA(369)//DEMIR	TCI091254	-0SE-0E-8DYR-0E-1YA-0YM	TCI	161897	
125	UKR-OD 761.93/AE.SQUARROSA(392)	CAWW04GH00022S	-0GH-0SE-030E-15E-0E-2YA-0YM	MX-TCI	161688	
126	PANDUR/AE.SQUARROSA(223)	CAWW04GH00079S	-0GH-0SE-030E-17E-0E-3YA-0YM	MX-TCI	161845	

## APPENDIX II. Crossing scheme of 124-winter synthetic hexaploid wheat under study.

Durum Parent	<i>Ae. taushii</i> parent										
	223	310	369	392	409	458	511	629	1026	1027	1031
AISBERG	-	-	X <sup>a</sup> (9) <sup>b</sup>	-	-	-	X (4) <sup>+c</sup> 1	-	-	-	-
LEUC 84693	-	-	-	-	X (4)	-	-	-	-	-	-
UKR-OD 761.93	-	-	-	X (4)	-	-	-	-	-	-	-
UKR-OD 952.92	-	-	-	-	-	-	-	-	-	-	X (8)
UKR-OD 1530.94	-	X (3)	-	X (3)	-	X (6)	-	X (4)+3	-	X (12)+2	-
PANDUR	X (15)	-	-	-	X (8)	-	-	-	-	-	-

<sup>a</sup>X: Original  
crosses

<sup>b</sup>( ): Number of entries from the same cross

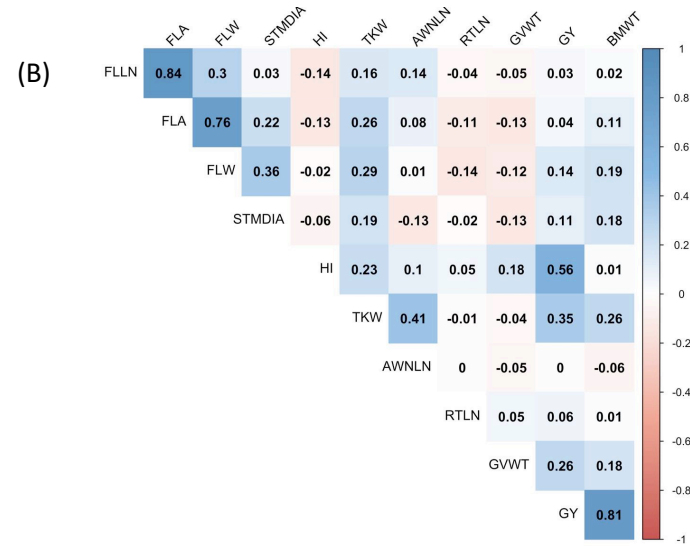
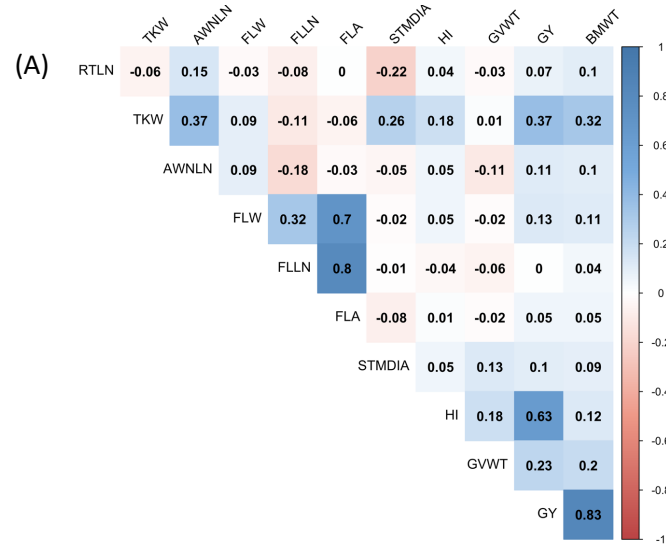
<sup>c</sup>+: Number of entries segregated and analyzed as unique entry based on phenotypic observation and kinship relationship value

**APPENDIX III.** Combined analysis of variance with means squares and broad sense heritability ( $H^2$ ) of 123 drought stressed synthetic hexaploid wheat grown in two seasons (2016 and 2017) in Konya, Turkey.

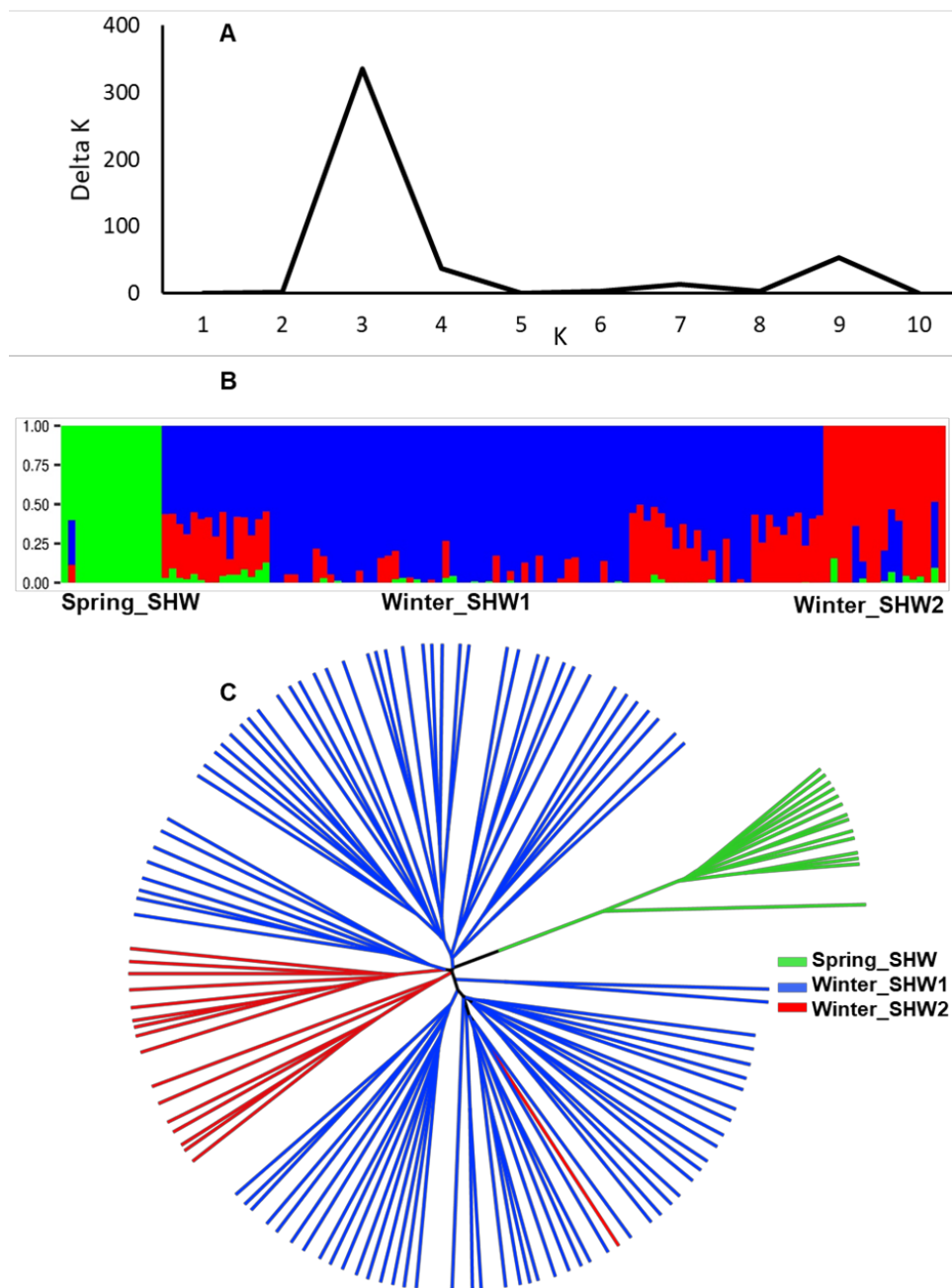
Source	Grain yield	Harvest index	Biomass weight	Thousand kernel weight	Grain volume weight	Awn length	Flag leaf length	Flag leaf width	Flag leaf area	Stem diameter	Root length
Year (YR )	9982	0.0073	229461	411***	84.3***	0.3	399***	0.0005	264***	0.04	263427***
Genotype(G)	5967*	0.0038*	27184*	27***	18.6**	5.8***	8*	0.0276***	14*	0.2***	9790*
G X YR	4376*	0.0027***	20126*	8***	10.9***	1.4***	7.7**	0.0152	12**	0.11	8510*
$H^2$	0.40	0.62	0.56	0.74	0.75	0.81	0.49	0.56	0.50	0.62	0.27

\*, \*\*, and \*\*\* Significant at the 0.05, 0.01, and 0.001 probability level, respectively.

**APPENDIX IV.** Pearson's correlation of grain yield and related traits of 123 synthetic hexaploid wheat based on adjusted best linear unbiased predictors in 2016 (A) and 2017 (B) growing season in Konya, Turkey.

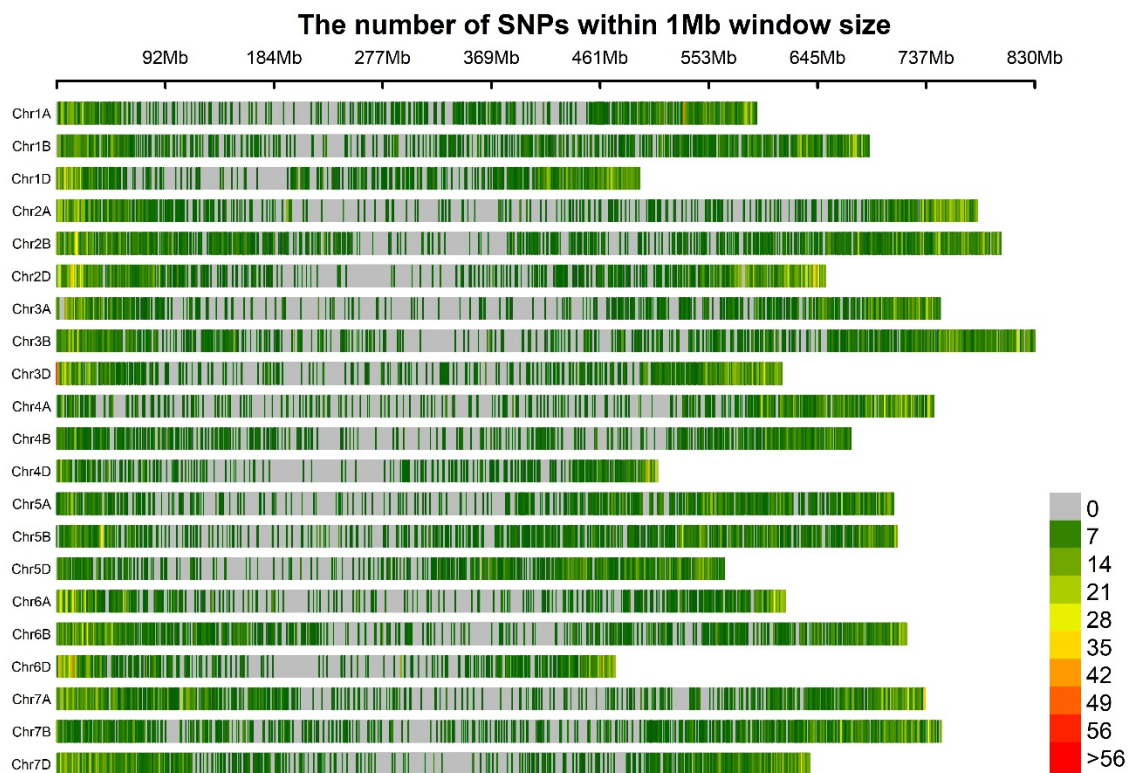


GY, Grain yield; HI, Harvest index; BMWT, Biomass weight; TKW, Thousand kernel weight, GVWT, Grain volume weight; AWNLN, Awn length; FLLN, Flag leaf length; FLW, Flag leaf width; FLA, Flag leaf area; STMDIA, Stem diameter; RTLN, Root length



APPENDIX V. Population structure analysis of 123 synthetic hexaploid wheat.

(A) Line graph of  $\Delta K$  over  $K$  from 1 to 10, and the highest peak was observed  $\Delta K=3$ , suggesting SHW has three subgroups. (B) Three subgroups (Spring\_SHW, Winter\_SHW1, and Winter\_SHW2) were identified from the STRUCTURE. (C) UPGMA cluster analysis



**APPENDIX VI.** Physical distribution of genotyping-by-sequencing derived SNPs within 1 Mb window size on 21 chromosomes of 123 synthetic hexaploid wheat.

**APPENDIX VII.** Details of significant markers associated with 11 traits in 123 synthetic hexaploid wheat grown in two seasons (2016 and 2017) in Konya, Turkey

Year	Trait	SNP	PVALUE	Allele	Effect	PVE (%)	Gene-ID	Human readable description
2016	Awn length	S2A_73179800	5.71	T/C	-0.54	2.01	TraesCS2A01G124000	Pentatricopeptide repeat-containing protein
2016	Awn length	S4A_602910768	6.07	A/G	0.48	2.89	-	-
2016	Awn length	S4A_691270332	5.80	T/G	-0.59	5.1	-	-
2016	Awn length	S4D_461573496	5.71	T/C	0.32	9.01	TraesCS4D01G290700.1	60S ribosomal protein L18a
2016	Awn length	S4D_490975981	4.56	C/T	0.47	4.4	TraesCS4D01G333500.1	Purple acid phosphatase
2016	Awn length	S5A_14299464	4.25	C/T	0.25	6.06	TraesCS5A01G018300	Thaumatococcus-like protein
2016	Awn length	S5A_562540562	11.67	C/T	-1.71	11.28	TraesCS5A01G361300.1	Guanine nucleotide exchange family protein
2016	Awn length	S5A_663583007	7.31	A/C	-0.72	10.69	TraesCS5A01G495900	ABC transporter B family protein
2016	Awn length	S5B_16679831	8.77	A/G	-0.78	17.64	-	-
2017	Awn length	S1D_374037078	6.26	G/T	0.67	1.19	TraesCS1D01G277100	DNA glycosylase
2017	Awn length	S2B_707465014	5.92	A/G	-0.32	12.86	-	-
2017	Awn length	S3B_511571190	4.74	C/T	0.59	1.06	-	-
2017	Awn length	S4B_654126249	7.99	G/A	-0.71	13.96	-	-
2017	Awn length	S5A_562786305	10.69	A/G	-1.59	8.12	-	-
2017	Awn length	S5A_680705502	10.10	C/A	0.92	20.05	TraesCS5A01G518600	-
2017	Awn length	S5B_43896804	7.31	C/T	-1.13	5.99	TraesCS5B01G038700- TraesCS5B01G038800	F-box family protein
2017	Awn length	S5D_431539773	4.33	G/A	0.28	5.76	-	-
2017	Awn length	S5D_434352162	4.23	A/T	0.3	8.14	-	-
2017	Awn length	S6B_643657	10.55	C/T	-1.04	5.66	TraesCS6B01G000900- TraesCS6B01G001000	Pollen Ole e 1 allergen/extensin and F-box protein
2017	Awn length	S7A_675273205	8.15	C/T	-0.46	3.11	TraesCS7A01G483800- TraesCS7A01G484000	Hydroxyproline-rich glycoprotein-like and Aspartic proteinase nepenthesin
2016	Biomass weight	S1D_441309135	4.82	G/C	-105.34	14.39	TraesCS1D01G357500.1	Protein DETOXIFICATION
2016	Biomass weight	S7B_283360494	4.19	G/A	25.62	9.34	-	-

2016	Biomass weight	S7B 450630784	4.06	G/A	-25.88	10.73	TraesCS7B01G242600.1	F-box family protein
2016	Biomass weight	S7B 604038670	4.26	C/G	23.45	8.63	-	
2017	Biomass weight	S2B 190232037	5.10	C/A	-37.57	2.66	TraesCS2B01G209000	Saccharopine dehydrogenase
2017	Biomass weight	S2B 691513138	4.66	T/C	-55.49	5.53	TraesCS2B01G494500	
2017	Biomass weight	S3A 24973175	4.21	T/C	-40.20	7.81	TraesCS3A01G047100	Aspartic proteinase nepenthesin-2
2017	Biomass weight	S3A 24973182	4.21	T/C	-40.20	7.81	TraesCS3A01G047100	Aspartic proteinase nepenthesin-2
2017	Biomass weight	S3A 25012018	6.08	G/A	-59.44	14.40	TraesCS3A01G047200- TraesCS3A01G047300	DNA-directed RNA polymerase subunit beta' and F-box domain containing protein
2017	Biomass weight	S4A 704559254	4.17	C/T	72.75	7.26	-	
2017	Biomass weight	S6D 459750797	4.53	C/G	33.69	4.90	TraesCS6D01G377000- TraesCS6D01G377100	Ethylene-responsive transcription factor and Myosin
2017	Biomass weight	S6D 459750812	4.53	G/A	-33.69	4.90	TraesCS6D01G377000- TraesCS6D01G377100	Ethylene-responsive transcription factor and Myosin
2017	Biomass weight	S6D 459750827	4.53	G/C	-33.69	4.90	TraesCS6D01G377000- TraesCS6D01G377100	Ethylene-responsive transcription factor and Myosin
2017	Biomass weight	S7B 682886226	4.26	C/G	42.83	3.11	TraesCS7B01G415000	Chaperone protein dnaJ
2017	Biomass weight	S7B 706522850	4.69	C/T	41.80	14.39	-	
2016	Flag leaf area	S1D 278097355	4.74	C/G	0.21	11.45	TraesCS1D01G197200.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
2016	Flag leaf area	S6B 120860110	4.01	A/G	0.17	9.25	TraesCS6B01G125800- TraesCS6B01G125900	Cytochrome P450 family protein, expressed
2016	Flag leaf area	S6B 120860130	4.01	T/A	-0.17	9.25	TraesCS6B01G125800- TraesCS6B01G125900	Ethylene-responsive transcription factor and Cytochrome P450 family protein, expressed
2017	Flag leaf area	S1A 516732460	6.90	G/A	-0.42	7.95	TraesCS1A01G326700.1	Citrate-binding protein
2017	Flag leaf area	S1A 532786451	4.23	G/T	-0.29	8.10	TraesCS1A01G345900- TraesCS1A01G346000	Splicing factor 3B subunit 5 and 30S ribosomal protein S15
2017	Flag leaf area	S1A 532786462	4.23	G/A	0.29	8.10	TraesCS1A01G345900- TraesCS1A01G346000	Splicing factor 3B subunit 5 and 30S ribosomal protein S15
2017	Flag leaf area	S1A 575597761	9.33	G/A	0.62	7.23	TraesCS1A01G418000.1	Polygalacturonase
2017	Flag leaf area	S1B 58989138	5.89	A/G	-0.32	4.42	TraesCS1B01G076100- TraesCS1B01G076200	Receptor-like protein kinase and Myb/SANT-like DNA-binding domain protein
2017	Flag leaf area	S2A 764065400	4.18	G/T	-0.19	3.75	TraesCS2A01G563200	NBS-LRR resistance-like protein
2017	Flag leaf area	S2A 29874199	17.79	G/A	4.65	23.07	-	-
2017	Flag leaf area	S2D 35564010	4.71	G/A	-0.41	7.12	-	-



2017	Flag leaf area	S4D 54054104	4.89	G/A	-0.28	8.07	TraesCS4D01G080000.1	Glutathione S-transferase
2017	Flag leaf area	S5A 587423540	6.15	G/C	0.26	3.90	TraesCS5A01G391700	MADS box transcription factor
2017	Flag leaf area	S6B 643131336	4.83	T/C	0.21	4.14	TraesCS6B01G369100	MADS-box transcription factor
2017	Flag leaf area	S6B 674558588	9.73	T/A	-0.65	10.30	TraesCS6B01G398900- TraesCS6B01G399000	Fructose-bisphosphate aldolase and Shugoshin-1
2017	Flag leaf area	S7D 558932149	7.71	C/A	0.50	8.00	TraesCS7D01G439100- TraesCS7D01G439200	SKP1-like protein and Gibberellin 2-oxidase
2017	Flag leaf area	S7D 10009696	4.63	G/C	-0.35	3.09	-	-
2017	Flag leaf area	S7D 638535043	4.00	G/A	0.26	3.09	-	-
2017	Flag leaf area	S7D 638535044	4.00	A/C	-0.26	3.09	-	-
2017	Flag leaf area	S7D 638535045	4.00	A/G	-0.26	3.09	-	-
2016	Flag leaf length	S1B 667135914	4.38	C/T	-0.16	20.75	TraesCS1B01G447400- TraesCS1B01G447500	Disease resistance protein RPM1 and Peroxisomal membrane protein PEX14
2016	Flag leaf length	S6D 1771825	4.65	A/C	-0.16	17.34	TraesCS6D01G003200	Amino acid transporter
2017	Flag leaf length	S1B 62791605	4.48	T/C	-0.41	6.58	TraesCS1B01G080400- TraesCS1B01G080500	Glycosyltransferase and Tetraspanin
2017	Flag leaf length	S1B 631203243	5.26	A/G	-0.21	9.81	TraesCS1B01G400600.1	Rp1-like protein
2017	Flag leaf length	S1D 382219667	6.72	C/G	-0.48	4.41	TraesCS1D01G283900.1	Chitinase
2017	Flag leaf length	S2A 29874199	22.52	G/A	4.74	32.27	-	-
2017	Flag leaf length	S2B 140752747	4.15	C/G	0.29	1.58	TraesCS2B01G167500.1	Cytochrome P450, putative
2017	Flag leaf length	S2D 642055122	4.12	C/T	0.25	5.86	TraesCS2D01G579800- TraesCS2D01G579900	protein kinase family protein and PLAC8 family protein
2017	Flag leaf length	S2D 71578532	4.68	G/A	-0.37	3.58	TraesCS2D01G122500.1	Plasma membrane fusion protein PRM1
2017	Flag leaf length	S4A 612662321	5.30	T/C	-0.23	5.31	TraesCS4A01G325200- TraesCS4A01G325300	F-box family protein
2017	Flag leaf length	S6D 463762312	5.72	A/G	0.25	5.99	TraesCS6D01G386200- TraesCS6D01G386300	Serine/threonine-protein kinase and Cytochrome P450, putative
2017	Flag leaf length	S7B 68562846	5.11	C/T	0.22	5.58	TraesCS7B01G063500- TraesCS7B01G063600	Phosphoinositide phosphatase family protein and Auxin responsive SAUR protein
2017	Flag leaf length	S7B 520419132	4.40	G/T	0.21	4.70	-	-
2016	Flag leaf width	S1B 453278609	4.21	C/T	-0.01	13.62	TraesCS1B01G257400- TraesCS1B01G257500	pale cress protein (PAC) and Coatomer subunit delta
2016	Flag leaf width	S1B 554003233	4.59	G/A	-0.02	9.34	TraesCS1B01G327900- TraesCS1B01G328000	Werner syndrome-like exonuclease, putative and U3 small nucleolar RNA-associated protein 15-like protein

2016	Flag leaf width	S6B_29539690	4.02	C/A	-0.02	11.53	TraesCS6B01G049800- TraesCS6B01G049900	-
2016	Flag leaf width	S6B_220551194	4.19	T/C	-0.03	13.81	TraesCS6B01G189600- TraesCS6B01G189700	DNA damage-inducible protein 1 and Protease do-like 1, chloroplastic
2016	Flag leaf width	S6B_320552308	4.67	A/G	-0.02	15.19	-	-
2016	Flag leaf width	S6D_16376439	4.85	T/C	-0.02	13.34	TraesCS6D01G040100.1	Mitochondrial transcription termination factor-like
2017	Flag leaf width	S1A_516732460	6.99	G/A	-0.03	9.46	TraesCS1A01G326700.1	Citrate-binding protein
2017	Flag leaf width	S1D_16816400	7.47	A/G	-0.03	9.86	TraesCS1D01G036000- TraesCS1D01G036100	Lecithin-cholesterol acyltransferase-like 1 and Protein FRIGIDA
2017	Flag leaf width	S2B_16009609	5.05	G/A	0.03	2.46	-	-
2017	Flag leaf width	S2B_48030550	5.11	T/C	-0.03	4.22	-	-
2017	Flag leaf width	S2D_32992152	11.31	G/T	-0.04	11.01	TraesCS2D01G077400.1	Actin cross-linking protein, putative (DUF569)
2017	Flag leaf width	S4B_534722043	5.75	G/A	0.04	9.41	TraesCS4B01G263700.1	OSBP (oxysterol binding protein)-related protein 1C
2017	Flag leaf width	S6B_26200560	7.13	C/A	0.03	12.29	TraesCS6B01G042800	F-box family protein
2017	Flag leaf width	S6B_73535204	5.34	C/A	0.03	5.51	TraesCS6B01G097100- TraesCS6B01G097200	Wall-associated receptor kinase 3 and Receptor-like protein kinase
2017	Flag leaf width	S6B_119525401	6.34	G/A	-0.03	1.86	TraesCS6B01G124500	BRCT domain-containing protein
2017	Flag leaf width	S6B_677338037	5.66	C/A	-0.02	1.6	TraesCS6B01G401300	Serine-rich protein
2016	Grain volulme weight	S2D_617414673	4.74	A/C	-3.20	11.76	-	-
2016	Grain volulme weight	S5A_423673926	5.21	A/G	-4.41	13.16	-	-
2016	Grain volulme weight	S6A_615815033	4.09	A/G	-2.87	10.5	TraesCS6A01G417500	Protein furry homolog-like protein
2016	Grain volulme weight	S7A_30902570	4.26	C/T	1.39	9.09	TraesCS7A01G062200- TraesCS7A01G062300	DNA topoisomerase
2016	Grain volulme weight	S7A_691163940	4.07	G/A	1.88	9.26	-	-
2017	Grain volulme weight	S1A_298646355	5.46	A/G	-0.47	10.50	TraesCS1A01G166200.1	Plant basic secretory family protein
2017	Grain volulme weight	S1A_522189599	4.11	G/A	-0.55	2.50	TraesCS1A01G334800	Cytochrome P450
2017	Grain volulme weight	S2A_758448348	5.63	A/G	0.50	14.71	TraesCS2A01G552200.1	Pyruvate decarboxylase
2017	Grain volulme weight	S2B_47837996	6.24	G/A	-0.49	1.26	TraesCS2B01G086200	Aspartic proteinase nepenthesin-1

2017	Grain volume weight	S3A_610441472	4.91	C/T	-0.46	8.79	-	-
2017	Grain volume weight	S4A_7441672	4.09	C/T	-0.80	16.17	TraesCS4A01G013100- TraesCS4A01G013200	Pyridoxal 5'-phosphate synthase subunit PdxT
2017	Grain volume weight	S4A_73454791	5.64	C/T	-0.63	5.45	TraesCS4A01G074200.2	Microtubule associated protein family protein, putative, expressed
2017	Grain volume weight	S7A_14787746	6.11	G/A	-0.86	10.32	-	-
2016	Grain yield	S1B_631118054	4.44	C/T	24.38	11.35	-	-
2016	Grain yield	S3A_686179591	4.08	G/A	-14.28	10.71	TraesCS3A01G445100	F-box family protein
2016	Grain yield	S7A_122984835	4.22	C/G	18.06	9.51	TraesCS7A01G167600- TraesCS7A01G167700	Growth-regulating factor and Myb/SANT-like DNA-binding domain protein
2016	Grain yield	S7A_97410463	4.15	T/C	-16.44	11.61	-	-
2016	Grain yield	S7A_110655838	5.38	A/G	19.97	13	-	-
2016	Grain yield	S7A_110655882	5.38	C/T	19.97	13	-	-
2016	Grain yield	S7A_110891713	4.41	G/C	-18.42	16.55	-	-
2016	Grain yield	S7A_110891755	4.41	C/T	18.42	16.55	-	-
2016	Grain yield	S7A_111089350	5.69	T/G	-19.60	16.36	-	-
2016	Grain yield	S7A_111089373	5.69	A/G	19.60	16.36	-	-
2016	Grain yield	S7A_112439457	4.51	G/A	-18.38	11.73	-	-
2016	Grain yield	S7A_112439468	4.51	C/G	18.38	11.73	-	-
2016	Grain yield	S7A_112439502	4.51	G/A	-18.38	11.73	-	-
2016	Grain yield	S7A_112439538	5.38	T/C	-19.97	13.27	-	-
2016	Grain yield	S7A_112977027	5.24	T/A	-19.17	12.79	TraesCS7A01G158200.1	Sentrin-specific protease
2016	Grain yield	S7A_113791836	4.85	C/G	19.97	12.76	-	-
2016	Grain yield	S7A_113791855	4.85	C/T	19.97	12.76	-	-
2016	Grain yield	S7A_114998636	4.53	G/A	-17.80	15.21	-	-
2016	Grain yield	S7A_115267913	5.22	A/G	18.88	17.32	-	-
2016	Grain yield	S7A_115313308	5.03	C/T	21.19	15.81	-	-
2016	Grain yield	S7B_717371800	4.25	C/T	-14.77	10.2	TraesCS7B01G459600- TraesCS7B01G459700	receptor kinase 1

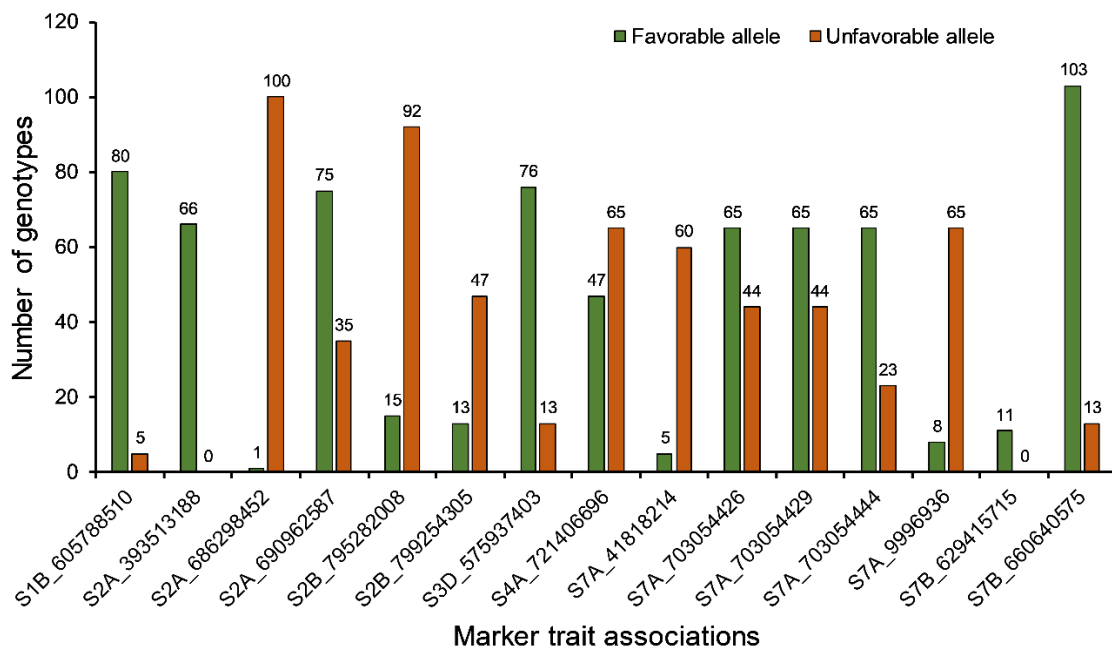
2017	Grain yield	S1B_5001751	4.06	C/A	-36.30	7.63	-	-
2017	Grain yield	S2B_67787236	4.37	A/G	-18.14	15.95	TraesCS2B01G106700- TraesCS2B01G106800	Wound-induced protein 1
2017	Grain yield	S3A_24993796	4.03	G/A	-15.29	10.11	TraesCS3A01G047200	DNA-directed RNA polymerase subunit beta'
2017	Grain yield	S3A_24993797	4.03	T/C	-15.29	10.11	TraesCS3A01G047200	DNA-directed RNA polymerase subunit beta'
2017	Grain yield	S3A_25012018	4.81	G/A	-20.02	12.74	TraesCS3A01G047200- TraesCS3A01G047300	DNA-directed RNA polymerase subunit beta' and F- box domain containing protein
2017	Grain yield	S3D_1203058	4.12	G/T	14.32	12.76	TraesCS3D01G002600- TraesCS3D01G002700	Aspartic proteinase nepenthesin-1 and Disease resistance protein RPM1
2017	Grain yield	S5B_598463062	5.36	T/C	-27.74	15.93	-	-
2017	Grain yield	S7A_27608184	5.48	G/T	-15.96	17.86	TraesCS7A01G057400	Glycosyltransferase
2016	Harvest Index	S1D_237334753	6.47	T/C	0.01	3.84		
2016	Harvest Index	S2A_747610082	5.21	G/A	-0.02	4.32	TraesCS2A01G528600.1	Heterogeneous nuclear ribonucleoprotein U-like protein 1
2016	Harvest Index	S2D_644672229	4.12	G/T	0.02	14.62		
2016	Harvest Index	S3A_593313534	13.56	T/C	0.08	16	TraesCS3A01G343700- TraesCS3A01G343800	WRKY transcription factor and Photosystem I reaction center subunit VIII
2016	Harvest Index	S3D_447198038	7.64	G/A	0.02	2.36		
2016	Harvest Index	S5B_598463062	8.85	T/C	-0.03	18.73		
2016	Harvest Index	S6B_555296159	6.21	C/T	-0.05	7.4	TraesCS6B01G309900.1	Cobyrinic acid synthase
2016	Harvest Index	S6D_157451060	4.01	G/A	-0.03	6.15	TraesCS6D01G170900.1	Cytochrome P450, putative
2016	Harvest Index	S6D_462272376	12.01	A/G	0.02	14.51	TraesCS6D01G382600.1	LOB domain protein-like
2016	Harvest Index	S7B_720164172	8.45	G/C	0.02	2.22	TraesCS7B01G463400	Non-lysosomal glucosylceramidase
2016	Root length	S3B_757480752	5.24	T/C	-0.38	18.51	TraesCS3B01G514800- TraesCS3B01G514900	Eukaryotic aspartyl protease family protein and Eukaryotic aspartyl protease family protein
2016	Root length	S5B_669373985	4.62	C/T	0.27	6.87	TraesCS5B01G502100- TraesCS5B01G502200	Dual specificity protein phosphatase and GRAM domain-containing protein / ABA-responsive
2016	Root length	S5B_669374027	4.62	C/T	0.27	6.87	TraesCS5B01G502100- TraesCS5B01G502200	Dual specificity protein phosphatase and GRAM domain-containing protein / ABA-responsive
2016	Root length	S6D_241296319	4.83	T/C	0.26	9.46	TraesCS6D01G185700- TraesCS6D01G185800	Nucleosome assembly protein 1-like 1
2016	Root length	S6D_431108774	4.01	G/A	-0.27	5.76	TraesCS6D01G332800.1	Protein DETOXIFICATION
2016	Root length	S6D_431173308	4.40	A/C	0.26	6.34	-	-

2016	Root length	S6D_435300571	5.09	A/G	0.49	16.7	-	-
2016	Root length	S6D_435306317	5.20	T/A	-0.31	8.73	-	-
2016	Root length	S6D_435306318	5.20	T/C	-0.31	8.73	-	-
2016	Root length	S6D_445773103	4.20	A/G	0.27	5.25	-	-
2016	Root length	S7A_94404310	4.01	G/A	0.52	7.5	TraesCS7A01G143200.2	Phosphatase 2C family protein
2017	Root length	S2B_84424899	4.10	G/A	128.42	10.09	-	-
2017	Root length	S2D_620326979	4.22	T/C	192.21	9.90	TraesCS2D01G541000.1	Disease resistance protein RPM1
2017	Root length	S6A_169248262	4.16	G/A	211.19	10.65	TraesCS6A01G166400- TraesCS6A01G166500	Glycosyltransferase family 92 protein and Non- structural maintenance of chromosome element 4
2017	Root length	S6A_169248303	4.16	A/G	-211.19	10.65	TraesCS6A01G166400- TraesCS6A01G166500	Glycosyltransferase family 92 protein and Non- structural maintenance of chromosome element 4
2016	Stem diameter	S1D_431523575	6.58	G/A	-0.06	10.31	TraesCS1D01G341500	Disease resistance protein (NBS-LRR class) family
2016	Stem diameter	S2B_153613233	6.40	C/T	0.08	11.03	TraesCS2B01G178500	Protein LTV1 like
2016	Stem diameter	S2D_38781826	8.28	G/A	-0.10	6.07	-	-
2016	Stem diameter	S3B_8490506	4.52	A/G	-0.12	2.15	TraesCS3B01G019600	E3 ubiquitin-protein ligase
2016	Stem diameter	S3B_64310629	4.16	G/C	-0.04	1.27	TraesCS3B01G095900	Major facilitator superfamily protein
2016	Stem diameter	S3B_64310630	4.16	T/C	-0.04	1.27	TraesCS3B01G095900	Major facilitator superfamily protein
2016	Stem diameter	S3B_647336163	5.03	C/T	-0.06	4.12	-	-
2016	Stem diameter	S3D_10133372	9.83	T/G	-0.11	8.63	TraesCS3D01G028500.1	Leucine-rich repeat receptor-like protein kinase family protein
2016	Stem diameter	S3D_13585155	7.28	G/C	-0.05	7.76	TraesCS3D01G037500.1	DUF1666 family protein
2016	Stem diameter	S5B_385190372	12.64	A/T	0.18	17.79	-	-
2016	Stem diameter	S6A_94238211	6.90	G/T	0.06	7.47	TraesCS6A01G122200.1	Protein kinase, putative
2016	Stem diameter	S7A_723295818	5.80	G/C	0.05	2.71	TraesCS7A01G549300	1,4-alpha-glucan branching enzyme
2017	Stem diameter	S1A_536172046	8.62	G/C	-0.07	9.45	-	-
2017	Stem diameter	S3A_26820384	4.43	T/C	0.05	2.70	TraesCS3A01G050400- TraesCS3A01G050500	Histone-lysine N-methyltransferase and GMP synthase [glutamine-hydrolyzing]
2017	Stem diameter	S4D_82741469	5.00	C/T	0.05	12.52	-	-
2017	Stem diameter	S4D_490591063	5.97	A/G	-0.15	2.37	-	-

2017	Stem diameter	S5A_539482288	6.71	C/A	0.14	3.39	-	-
2017	Stem diameter	S6A_611857844	6.97	G/C	-0.08	9.07	TraesCS6A01G406700	NAC domain protein,
2017	Stem diameter	S6B_610963076	5.70	G/T	0.06	7.68	TraesCS6B01G346900- TraesCS6B01G347000	NBS-LRR disease resistance protein and F-box protein-like
2017	Stem diameter	S6D_472404065	6.46	C/T	0.06	4.19	TraesCS6D01G404800- TraesCS6D01G404900	Calcium-dependent protein kinase and Protein SAWADEE HOMEODOMAIN-like protein 2
2017	Stem diameter	S6D_94986698	5.76	G/T	0.10	18.08	-	-
2017	Stem diameter	S7A_65090371	7.19	G/C	-0.10	3.88	TraesCS7A01G107600- TraesCS7A01G107700	Mediator complex, subunit Med7 and 3-ketoacyl-CoA synthase
2017	Stem diameter	S7B_159873337	4.03	C/T	0.06	28.77	TraesCS7B01G132100- TraesCS7B01G132200	Loricrin-like and 4-hydroxy-tetrahydrodipicolinate synthase
2016	Thousand kernel weight	S2A_47781717	4.52	A/G	0.93	4.16	TraesCS2A01G093400- TraesCS2A01G093500	Myb/SANT-like DNA-binding domain protein and F-box family protein
2016	Thousand kernel weight	S2B_409327412	5.45	C/A	-1.28	5.12	TraesCS2B01G293500	Tetratricopeptide repeat (TPR)-like superfamily protein
2016	Thousand kernel weight	S2D_7309581	6.09	G/T	0.70	14.72	-	-
2016	Thousand kernel weight	S3B_758391015	5.42	A/C	-0.81	9.73	TraesCS3B01G516300- TraesCS3B01G516400	ATP synthase subunit a
2016	Thousand kernel weight	S4A_625466381	4.12	G/T	1.21	15.3	TraesCS4A01G347500- TraesCS4A01G347600	Receptor-like protein kinase and protein kinase family protein
2016	Thousand kernel weight	S4A_402470267	6.97	C/T	1.65	18.64	-	-
2016	Thousand kernel weight	S4D_509427923	4.91	G/C	-1.72	10.06	TraesCS4D01G364700	Cytochrome P450 family protein
2016	Thousand kernel weight	S6D_452410667	8.16	A/G	-1.54	17.66	TraesCS6D01G360800	protein kinase family protein
2016	Thousand kernel weight	S7B_714380330	5.06	T/C	-0.95	11.46	-	-
2017	Thousand kernel weight	S1A_17748135	5.00	C/G	0.62	8.59	-	-
2017	Thousand kernel weight	S2B_683472157	4.31	C/T	-1.48	12.74	-	-
2017	Thousand kernel weight	S2D_7309581	7.60	G/T	0.83	8.51	-	-
2017	Thousand kernel weight	S3A_52455011	6.02	T/C	1.31	1.64	TraesCS3A01G081100	Major facilitator superfamily protein
2017	Thousand kernel weight	S4A_681180933	6.12	T/G	-0.81	14.08	TraesCS4A01G408300	Xyloglucan galactosyltransferase KATAMARI1-like protein
2017	Thousand kernel weight	S4B_11905230	8.94	C/G	-1.11	3.91	TraesCS4B01G016200.1	LOB domain-containing protein, putative

2017	Thousand kernel weight	S4B 407743714	15.99	A/T	2.31	13.38	TraesCS4B01G187300- TraesCS4B01G187400	Protein COBRA, putative and Hexosyltransferase
2017	Thousand kernel weight	S4B 637722874	5.17	C/T	0.86	2.03	TraesCS4B01G344200.1	Zinc finger (C3HC4-type RING finger) family protein
2017	Thousand kernel weight	S5B 288308122	4.31	C/G	-0.65	2.21	TraesCS5B01G156100	Alpha/beta-Hydrolases superfamily protein
2017	Thousand kernel weight	S5B 616966405	6.91	G/A	1.29	22.21	-	-
2017	Thousand kernel weight	S7D 46755477	6.38	G/A	-0.76	4.29	-	-

PVE: Phenotypic variance explained



**APPENDIX VIII.** Number of synthetic hexaploid wheat germplasm having either favorable or unfavorable alleles associated with common bunt resistance.



**APPENDIX IX.** Soil sample analysis in 2016 and 2017 growing season in Konya, Turkey

Year	Sample	pH	Electrical conductivity	P <sub>2</sub> O <sub>5</sub>	K <sub>2</sub> O	Organic matter	CaCO <sub>3</sub>
			Mmhos cm <sup>-1</sup>	Kg ha <sup>-1</sup>	Kg ha <sup>-1</sup>	%	%
2016	1	7.70	1.21	2.071	17.19	2.08	30.19
2016	2	7.74	1.15	2.019	17.19	1.98	30.74
Mean		7.72	1.18	2.045	17.19	2.03	30.465
2017	1	8.21	0.71	0.888	28.57	1.30	31.57
2017	2	8.20	0.66	0.853	28.59	1.35	28.27
Mean		8.205	0.685	0.8705	28.58	1.33	29.92

**APPENDIX X.** Details of top ranking 13 synthetic hexaploid wheat and checks (Gerek and Karahan) based on two years combined data.

Entry#	Pedigree name	Ca	Cd	Co	Cu	Fe	Li	Mg	Mn	Ni	Zn	Favalle	GPC	GY
		mg/Kg	mg/Kg	mg/Kg	mg/Kg	mg/Kg	mg/Kg	mg/Kg	mg/Kg	mg/Kg	mg/Kg		g/Kg	g/m <sup>2</sup>
119	LANGDON/KU-2093	90.87	0.08	0.02	8.79	56.02	0.27	1707.43	58.17	0.79	29.57	27	162.54	212.97
81	PANDUR/AE.SQUARROSA(223)	108.75	0.08	0.04	8.73	55.67	0.60	1986.62	59.37	1.34	30.20	37	148.86	321.63
107	LANGDON/IG 126387	96.59	0.09	0.02	7.38	54.27	0.44	1673.35	49.30	0.61	30.24	24	151.83	221.48
83	PANDUR/AE.SQUARROSA(223)	81.16	0.08	0.04	9.80	54.22	0.38	1862.02	61.53	0.81	33.03	26	147.56	385.60
118	LANGDON/KU-20-9	80.98	0.08	0.03	8.46	53.73	0.29	1781.54	56.01	0.86	33.16	28	158.35	205.71
117	LANGDON/KU-2079	89.35	0.07	0.03	8.36	53.35	0.24	1751.60	59.37	0.79	34.47	26	167.43	260.41
109	LANGDON/KU-2074	92.33	0.09	0.04	7.89	51.37	0.27	1762.32	55.09	1.02	28.91	23	161.75	150.51
114	LANGDON/KU-2097	81.21	0.08	0.03	7.88	50.87	0.27	1568.09	53.68	0.81	31.43	25	162.61	213.78
12	AISBERG/AE.SQUARROSA(511)	96.75	0.08	0.03	9.15	50.38	0.26	1894.32	63.00	0.83	33.67	28	149.60	226.72
82	PANDUR/AE.SQUARROSA(223)	101.91	0.08	0.04	8.68	50.29	0.40	1892.40	55.87	1.07	32.58	34	149.80	300.11
125	UKR-OD 761.93/AE.SQUARROSA(392)	79.23	0.08	0.03	9.29	49.98	0.28	1793.75	56.75	0.73	32.58	27	137.24	323.52
90	PANDUR/AE.SQUARROSA(223)	99.33	0.08	0.03	9.09	49.70	0.40	1899.86	56.86	0.90	31.93	31	153.63	275.42
102	PANDUR/AE.SQUARROSA(409)	82.08	0.07	0.02	8.31	49.45	0.24	1722.92	52.75	0.70	33.74	28	156.68	181.83
127	Gerek	72.19	0.05	0.02	4.27	30.09	0.31	1179.2	34.01	0.55	16.51	-	133.21	301.1
128	Karahan	80.32	0.05	0.02	5.94	33.75	0.18	1314.79	33.72	0.45	18.23	-	141.35	300.11

Favallele, favorable allele; GPC, grain protein concentration; GY, grain yield

APPENDIX XI. Details of significant markers associated with 10 grain minerals from genome wide association study of 123 synthetic hexaploid wheats grown in 2016 and 2017 in Konya, Turkey.

Trait	Year	Dataset	SNP	Chromosome	Position (bp)	Alleles	Favorable allele	PVE (%)	SNP effect	Pvalue	-log10(P)
Zinc	combined	CBLUP	S1A_1846816	1A	1846816	A/G	A	3.0	-0.655	1.47E-07	6.83
Nickel	2016	BLUP16	S1A_402236557	1A	402236557	G/A	A	19.3	-0.153	8.00E-13	12.10
Iron	combined	CBLUP	S1A_584413238	1A	584413238	T/G	G	11.2	-3.354	1.17E-06	5.93
Iron	combined	CBLUP	S1A_584413248	1A	584413248	G/C	C	11.2	-3.354	1.17E-06	5.93
Cadmium	2017	BLUP17	S1A_587736663	1A	587736663	G/A	A	4.0	-0.002	4.25E-07	6.37
Copper	2016	BLUP16	S1B_544997339	1B	544997339	G/A	G	10.6	0.294	3.02E-07	6.52
Lithium	2017	BLUP17	S1B_606491241	1B	606491241	T/C	C	13.5	-0.039	6.29E-10	9.20
Lithium	combined	CBLUP	S1B_606491241	1B	606491241	T/C	C	12.3	-0.011	6.74E-11	10.17
Magnesium	2017	BLUP17	S1B_6867825	1B	6867825	C/A	A	1.4	-121.477	1.62E-08	7.79
Calcium	combined	CBLUP	S1B_6867825	1B	6867825	C/A	A	4.7	-2.938	1.08E-10	9.97
Magnesium	combined	CBLUP	S1B_6867825	1B	6867825	C/A	A	1.4	-73.196	5.35E-08	7.27
Lithium	combined	CBLUP	S1D_17169160	1D	17169160	C/G	C	5.1	-0.011	6.32E-07	6.20
Magnesium	combined	CBLUP	S1D_466362317	1D	466362317	C/T	T	9.6	64.345	1.13E-07	6.95
Lithium	combined	CBLUP	S2A_11222292	2A	11222292	C/G	G	15.2	0.018	2.23E-08	7.65
Cadmium	2017	BLUP17	S2A_23985636	2A	23985636	G/A	G	1.8	0.011	5.59E-14	13.25
Copper	combined	CBLUP	S2A_738732586	2A	738732586	T/G	G	3.1	-0.430	2.57E-07	6.59
Copper	2016	BLUP16	S2A_742969119	2A	742969119	A/G	G	5.3	0.578	1.18E-07	6.93
Zinc	2016	BLUP16	S2A_742969119	2A	742969119	A/G	G	8.9	1.344	1.19E-06	5.92
Zinc	2017	BLUP17	S2A_750621751	2A	750621751	C/A	C	10.3	1.538	2.10E-07	6.68
Calcium	2017	BLUP17	S2B_502127437	2B	502127437	C/T	C	9.3	-1.966	6.64E-07	6.18
Nickel	2016	BLUP16	S2D_48611294	2D	48611294	C/T	T	21.1	0.234	7.09E-14	13.15
Nickel	combined	CBLUP	S2D_48611294	2D	48611294	C/T	T	20.2	0.073	5.56E-10	9.25
Magnesium	2017	BLUP17	S2D_506778844	2D	506778844	T/G	G	11.8	-71.163	5.35E-09	8.27
Lithium	combined	CBLUP	S2D_517038749	2D	517038749	C/A	C	1.9	0.014	3.40E-07	6.47
Lithium	2017	BLUP17	S2D_572031650	2D	572031650	G/A	A	12.6	-0.049	5.23E-09	8.28

Lithium	combined	CBLUP	S2D_572031650	2D	572031650	G/A	A	14.6	-0.014	1.05E-10	9.98
Manganese	2016	BLUP16	S2D_58740285	2D	58740285	A/G	A	8.7	-2.742	9.39E-07	6.03
Manganese	combined	CBLUP	S2D_58740285	2D	58740285	A/G	A	10.9	-1.267	1.71E-08	7.77
Calcium	combined	CBLUP	S2D_631996199	2D	631996199	G/A	A	7.6	-1.686	2.74E-08	7.56
Cadmium	2017	BLUP17	S2D_80258448	2D	80258448	C/A	C	7.0	0.003	2.15E-07	6.67
Copper	2016	BLUP16	S3A_23297031	3A	23297031	T/C	C	5.8	-0.374	3.45E-07	6.46
Nickel	combined	CBLUP	S3A_530501108	3A	530501108	T/C	T	13.3	0.029	5.04E-07	6.30
Zinc	2017	BLUP17	S3A_534469328	3A	534469328	A/T	A	11.1	-1.094	4.19E-07	6.38
Magnesium	combined	CBLUP	S3A_534469328	3A	534469328	A/T	A	14.6	-31.191	2.62E-07	6.58
Manganese	combined	CBLUP	S3A_534469328	3A	534469328	A/T	A	14.3	-1.466	1.25E-08	7.90
Zinc	combined	CBLUP	S3A_534469328	3A	534469328	A/T	A	12.7	-0.843	5.52E-11	10.26
Iron	combined	CBLUP	S3A_534535579	3A	534535579	G/C	G	13.2	2.675	4.40E-07	6.36
Calcium	combined	CBLUP	S3A_593702925	3A	593702925	A/G	G	7.9	1.270	7.47E-08	7.13
Cobalt	2017	BLUP17	S3A_736119715	3A	736119715	A/C	C	18.5	0.001	1.39E-08	7.86
Cadmium	2017	BLUP17	S3A_742590055	3A	742590055	C/T	C	8.7	-0.003	2.02E-09	8.69
Zinc	combined	CBLUP	S3B_206308044	3B	206308044	G/A	A	1.8	-0.551	4.20E-07	6.38
Calcium	2017	BLUP17	S3B_548275272	3B	548275272	T/C	T	2.7	2.349	4.54E-07	6.34
Calcium	combined	CBLUP	S3B_655010350	3B	655010350	C/T	T	2.9	2.962	1.13E-07	6.95
Copper	combined	CBLUP	S3B_689167760	3B	689167760	C/T	T	9.2	0.281	2.68E-07	6.57
Magnesium	2017	BLUP17	S3B_727935439	3B	727935439	G/A	A	6.7	-59.089	9.76E-08	7.01
Zinc	2017	BLUP17	S3B_78136780	3B	78136780	T/C	T	1.8	0.995	1.32E-06	5.88
Zinc	2016	BLUP16	S3B_813450132	3B	813450132	C/A	C	4.7	2.725	2.30E-09	8.64
Calcium	2017	BLUP17	S3D_45073985	3D	45073985	T/A	A	21.5	-3.463	1.38E-07	6.86
Lithium	2017	BLUP17	S3D_610567350	3D	610567350	G/T	T	22.6	0.066	2.30E-09	8.64
Lithium	combined	CBLUP	S3D_610567350	3D	610567350	G/T	T	16.5	0.021	3.62E-10	9.44
Zinc	combined	CBLUP	S4A_681683160	4A	681683160	G/A	A	13.8	-0.782	7.93E-07	6.10
Magnesium	combined	CBLUP	S4A_740606543	4A	740606543	A/G	G	9.7	81.062	1.01E-07	7.00
Copper	2016	BLUP16	S4B_37424735	4B	37424735	C/T	C	5.2	-0.340	2.98E-07	6.53

Manganese	combined	CBLUP	S4B_407743657	4B	407743657	C/A	A	13.4	-2.502	1.56E-10	9.81
Zinc	combined	CBLUP	S4B_408606348	4B	408606348	G/C	C	14.1	-1.022	4.21E-08	7.38
Magnesium	combined	CBLUP	S4B_624138956	4B	624138956	C/T	C	1.8	-50.785	1.17E-06	5.93
Nickel	2017	BLUP17	S4D_381885629	4D	381885629	A/C	C	13.6	0.052	1.10E-06	5.96
Magnesium	combined	CBLUP	S4D_490394558	4D	490394558	T/G	G	7.9	-27.722	1.37E-06	5.86
Lithium	2017	BLUP17	S5A_135164381	5A	135164381	G/C	C	4.4	-0.040	9.27E-08	7.03
Lithium	combined	CBLUP	S5A_135164381	5A	135164381	G/C	C	7.2	-0.013	2.65E-09	8.58
Zinc	2016	BLUP16	S5A_395285802	5A	395285802	G/A	A	10.6	-0.736	6.78E-08	7.17
Zinc	combined	CBLUP	S5A_552354940	5A	552354940	C/G	G	8.8	1.664	6.79E-07	6.17
Copper	combined	CBLUP	S5A_643024354	5A	643024354	A/C	A	1.6	-0.275	1.02E-06	5.99
Magnesium	2016	BLUP16	S5B_405724949	5B	405724949	G/A	G	12.0	99.659	1.00E-06	6.00
Copper	2016	BLUP16	S5B_607870649	5B	607870649	T/C	C	14.4	-0.527	9.18E-10	9.04
Copper	combined	CBLUP	S5B_607870649	5B	607870649	T/C	C	17.1	-0.266	3.51E-07	6.46
Nickel	combined	CBLUP	S5B_616966405	5B	616966405	G/A	G	9.2	0.031	1.14E-06	5.94
Magnesium	2017	BLUP17	S5B_65069754	5B	65069754	C/G	C	2.4	-80.852	1.28E-06	5.89
Magnesium	2017	BLUP17	S5D_28880704	5D	28880704	C/G	C	3.8	-54.631	4.13E-08	7.38
Manganese	combined	CBLUP	S5D_29132492	5D	29132492	C/G	C	4.4	-1.391	3.25E-09	8.49
Copper	combined	CBLUP	S5D_493381192	5D	493381192	C/T	T	1.8	0.152	4.84E-07	6.31
Nickel	combined	CBLUP	S6A_430583367	6A	430583367	C/T	C	4.7	-0.090	1.81E-08	7.74
Calcium	2017	BLUP17	S6A_50345873	6A	50345873	C/T	C	9.0	-1.345	6.73E-08	7.17
Calcium	2017	BLUP17	S6A_592562315	6A	592562315	C/G	G	6.9	8.350	2.34E-11	10.63
Copper	combined	CBLUP	S6A_613579920	6A	613579920	C/G	C	1.2	-0.210	4.94E-08	7.31
Nickel	combined	CBLUP	S6A_99795469	6A	99795469	T/C	C	14.9	-0.039	3.17E-07	6.50
Calcium	2016	BLUP16	S6B_109760004	6B	109760004	T/A	T	19.9	2.763	3.77E-09	8.42
Copper	combined	CBLUP	S6B_27918199	6B	27918199	C/T	T	10.9	0.343	1.10E-10	9.96
Calcium	2016	BLUP16	S6B_32333184	6B	32333184	C/G	G	9.9	3.552	1.98E-09	8.70
Calcium	combined	CBLUP	S6B_32333184	6B	32333184	C/G	G	13.0	3.045	4.87E-11	10.31
Zinc	2016	BLUP16	S6B_482791655	6B	482791655	T/C	C	9.4	-1.410	8.46E-07	6.07

Manganese	2016	BLUP16	S6B_48536435	6B	48536435	A/T	T	11.5	5.491	1.08E-06	5.97
Calcium	2016	BLUP16	S6B_576856920	6B	576856920	C/G	G	9.3	2.298	9.07E-07	6.04
Copper	2016	BLUP16	S6B_6241996	6B	6241996	A/C	C	10.0	0.684	9.94E-11	10.00
Calcium	combined	CBLUP	S6B_658724336	6B	658724336	C/T	T	9.6	1.799	2.03E-08	7.69
Lithium	2017	BLUP17	S6D_30744756	6D	30744756	A/G	G	20.3	0.041	8.40E-10	9.08
Lithium	combined	CBLUP	S6D_30744756	6D	30744756	A/G	G	12.6	0.012	1.57E-11	10.80
Cadmium	combined	CBLUP	S6D_447907113	6D	447907113	T/C	T	14.4	0.001	9.57E-07	6.02
Cobalt	2017	BLUP17	S6D_452082847	6D	452082847	A/C	C	25.2	0.001	2.46E-07	6.61
Magnesium	2017	BLUP17	S7A_116286914	7A	116286914	T/G	G	7.3	-94.175	1.80E-07	6.74
Calcium	2016	BLUP16	S7A_34297426	7A	34297426	G/A	G	11.8	2.574	1.14E-07	6.94
Cobalt	2017	BLUP17	S7D_17753962	7D	17753962	A/T	T	21.2	0.001	1.30E-10	9.89

APPENDIX XII. List of significant markers associated with 10 grain minerals and gene annotation to test the reliability of the MTA from genome wide association study of 123 synthetic hexaploid wheats grown in 2016 and 2017 in Konya, Turkey.

Trait	Year	Dataset	SNPID	Gene Name	Gene Annotation
Cadmium	2017	BLUP17	S2D_80258448	TraesCS2D01G136700- TraesCS2D01G136800	Glycerol-3-phosphate acyltransferase 3, putative and Cysteine-rich receptor-kinase-like protein (Salt stress response/antifungal)
Calcium	2016	BLUP16	S6B_576856920	TraesCS6B01G327300	Allergen, putative
Calcium	2016	BLUP16	S7A_34297426	TraesCS7A01G068100- TraesCS7A01G068200	Transcription elongation factor 1 and NAC domain-containing protein, putative ( No apical meristem (NAM) protein)
Calcium	2016	BLUP16	S6B_109760004	TraesCS6B01G117600- TraesCS6B01G117700	Pyridoxal 5'-phosphate synthase subunit PdxT and ATP synthase gamma chain
Calcium	2016	BLUP16	S6B_32333184	TraesCS6B01G053100	MYB transcription factor
Calcium	2017	BLUP17	S6A_50345873	TraesCS6A01G081700- TraesCS6A01G081800	Peroxidase and Cytochrome P450
Calcium	combined	CBLUP	S2D_631996199	TraesCS2D01G559600- TraesCS2D01G559700	transmembrane protein, putative (DUF594) and Zn-dependent exopeptidases superfamily protein
Calcium	combined	CBLUP	S6B_658724336	TraesCS6B01G384300- TraesCS6B01G384400	Leucine-rich repeat receptor-like protein kinase family protein
Calcium	combined	CBLUP	S3B_655010350	TraesCS3B01G417900- TraesCS3B01G418000	Protein bps1, chloroplastic and Adenine nucleotide alpha hydrolases-like superfamily protein (Universal stress protein family)
Cobalt	2017	BLUP17	S6D_452082847	TraesCS6D01G360300	F-box family protein
Cobalt	2017	BLUP17	S3A_736119715	TraesCS3A01G518900- TraesCS3A01G519000	DNA topoisomerase and Cytochrome P450, putative
Copper	2016	BLUP16	S3A_23297031	TraesCS3A01G042900- TraesCS3A01G043000	DNA-binding storekeeper protein-related transcriptional regulator and 30S ribosomal protein S10

Copper	2016	BLUP16	S4B_37424735	TraesCS4B01G049200- TraesCS4B01G049300	50S ribosomal protein L32, chloroplastic and GPN-loop GTPase-like protein
Copper	2016	BLUP16	S2A_742969119	TraesCS2A01G519900- TraesCS2A01G520000	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
Copper	combined	CBLUP	S2A_738732586	TraesCS2A01G514200- TraesCS2A01G514300	AP2-like ethylene-responsive transcription factor and Ethylene-responsive transcription factor
Iron	combined	CBLUP	S1A_584413238	TraesCS1A01G432800- TraesCS1A01G432900	Metacaspase-1 and Na-translocating NADH-quinone reductase subunit A
Iron	combined	CBLUP	S1A_584413248	TraesCS1A01G432800- TraesCS1A01G432900	Metacaspase-1 and Na-translocating NADH-quinone reductase subunit A
Iron	combined	CBLUP	S3A_534535579	TraesCS3A01G300700	Serpin family protein
Lithium	2017	BLUP17	S2D_572031650	TraesCS2D01G466400	Leucine-rich repeat receptor-like protein [disease resistance protein (TIR-NBS-LRR class) family]
Lithium	2017	BLUP17	S5A_135164381	TraesCS5A01G096300	GDSL esterase/lipase
Lithium	2017	BLUP17	S1B_606491241	TraesCS1B01G375400	Kinase family protein
Lithium	2017	BLUP17	S6D_30744756	TraesCS6D01G064500- TraesCS6D01G064600	F-box protein and F-box SKIP23-like protein
Magnesium	2017	BLUP17	S3B_727935439	TraesCS3B01G479800- TraesCS3B01G479900	F-box family protein and CBS domain-containing protein-like
Magnesium	combined	CBLUP	S4B_624138956	TraesCS4B01G333400- TraesCS4B01G333500	carboxyl-terminal peptidase, putative (DUF239) and carboxyl-terminal peptidase (DUF239)
Magnesium	combined	CBLUP	S3A_534469328	TraesCS3A01G300400	Protein DETOXIFICATION
Magnesium	combined	CBLUP	S4D_490394558	TraesCS4D01G333000- TraesCS4D01G333100	ROP guanine nucleotide exchange factor 10 and F-box family protein



Magnesium	combined	CBLUP	S4A_740606543	TraesCS4A01G490700	Leucine-rich repeat receptor-like protein kinase family protein, putative
Manganese	2016	BLUP16	S2D_58740285	TraesCS2D01G106500- TraesCS2D01G106600	F-box protein and Potassium transporter
Manganese	2016	BLUP16	S6B_48536435	TraesCS6B01G071900	thionin-like protein
Manganese	combined	CBLUP	S4B_407743657	TraesCS4B01G187300- TraesCS4B01G187400	Protein COBRA, putative and Hexosyltransferase ( Glycosyl transferase family 8)
Nickel	2016	BLUP16	S1A_402236557	TraesCS1A01G229600	Elongation factor 4
Nickel	combined	CBLUP	S6A_430583367	TraesCS6A01G228400	Prolyl oligopeptidase family protein
Nickel	combined	CBLUP	S3A_530501108	TraesCS3A01G296300	LURP-one-like protein
Zinc	2016	BLUP16	S6B_482791655	TraesCS6B01G268400- TraesCS6B01G268500	FBD-associated F-box protein and Sentrin-specific protease 2
Zinc	2017	BLUP17	S3B_78136780	TraesCS3B01G111800- TraesCS3B01G111900	Accelerated cell death 11 and F-box protein
Zinc	2017	BLUP17	S2A_750621751	TraesCS2A01G536700	Flowering Locus T-like protein, putative
Zinc	combined	CBLUP	S4B_408606348	TraesCS4B01G187600	Chaperone protein dnaJ
Zinc	combined	CBLUP	S4A_681683160	TraesCS4A01G408900	Cytoplasmic FMR1-interacting
Zinc	combined	CBLUP	S1A_1846816	TraesCS1A01G003300- TraesCS1A01G003400	Protein ROOT HAIR DEFECTIVE 3 homolog
Zinc	combined	CBLUP	S3B_206308044	TraesCS3B01G192400- TraesCS3B01G192500	Phosphate translocator and Receptor-like kinase (Leucine rich repeat N-terminal domain)