

## Article

# GWAS in a Collection of Bulgarian Old and Modern Bread Wheat Accessions Uncovers Novel Genomic Loci for Grain Protein Content and Thousand Kernel Weight

Tania Kartseva <sup>1,†</sup>, Vladimir Aleksandrov <sup>1,†</sup> , Ahmad M. Alqudah <sup>2</sup> , Mian Abdur Rehman Arif <sup>3</sup> , Konstantina Kocheva <sup>1</sup> , Dilyana Doneva <sup>1</sup> , Katelina Prokopova <sup>1</sup>, Andreas Börner <sup>4</sup>  and Svetlana Misheva <sup>1,\*</sup> 

<sup>1</sup> Institute of Plant Physiology and Genetics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 21, 1113 Sofia, Bulgaria; tania\_kartseva@abv.bg (T.K.); aleksandrov@gbg.bg (V.A.)

<sup>2</sup> Biological Science Program, Department of Biological and Environmental Sciences, College of Art and Science, Qatar University, Doha P.O. Box 2713, Qatar; aalqudah@qu.edu.qa or ahqudah@gmail.com

<sup>3</sup> Nuclear Institute for Agriculture and Biology College, Pakistan Institute of Engineering and Applied Sciences (NIAB-C, PIEAS), Jhang Road, Faisalabad 38000, Pakistan; m.a.rehman.arif@gmail.com

<sup>4</sup> Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben), OT Gatersleben, Corrensstraße 3, 06466 Seeland, Germany; boerner@ipk-gatersleben.de

\* Correspondence: slandjeva@gmail.com

† These authors contributed equally to this work.

**Abstract:** Genetic enhancement of grain production and quality is a priority in wheat breeding projects. In this study, we assessed two key agronomic traits—grain protein content (GPC) and thousand kernel weight (TKW)—across 179 Bulgarian contemporary and historic varieties and landraces across three growing seasons. Significant phenotypic variation existed for both traits among genotypes and seasons, and no discernible difference was evident between the old and modern accessions. To understand the genetic basis of the traits, we conducted a genome-wide association study with MLM using phenotypic data from the crop seasons, best linear unbiased estimators, and genotypic data from the 25K Infinium iSelect array. As a result, we detected 16 quantitative trait nucleotides (QTNs) associated with GPC and 15 associated with TKW, all of which passed the false discovery rate threshold. Seven loci favorably influenced GPC, resulting in an increase of 1.4% to 8.1%, while four loci had a positive impact on TKW with increases ranging from 1.9% to 8.4%. While some loci confirmed previously published associations, four QTNs linked to GPC on chromosomes 2A, 7A, and 7B, as well as two QTNs related to TKW on chromosomes 1B and 6A, may represent novel associations. Annotations for proteins involved in the senescence-associated nutrient remobilization and in the following buildup of resources required for seed germination have been found for selected putative candidate genes. These include genes coding for storage proteins, cysteine proteases, cellulose-synthase, alpha-amylase, transcriptional regulators, and F-box and RWP-RK family proteins. Our findings highlight promising genomic regions for targeted breeding programs aimed at improving grain yield and protein content.

**Keywords:** association mapping; candidate genes; grain protein content; TKW; grain quality; grain yield; *Triticum aestivum* L.



**Citation:** Kartseva, T.; Aleksandrov, V.; Alqudah, A.M.; Arif, M.A.R.; Kocheva, K.; Doneva, D.; Prokopova, K.; Börner, A.; Misheva, S. GWAS in a Collection of Bulgarian Old and Modern Bread Wheat Accessions Uncovers Novel Genomic Loci for Grain Protein Content and Thousand Kernel Weight. *Plants* **2024**, *13*, 1084. <https://doi.org/10.3390/plants13081084>

Academic Editors: Jianghua Chen and Zhenlan Liu

Received: 16 February 2024

Revised: 3 April 2024

Accepted: 10 April 2024

Published: 12 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Wheat is a unique cereal crop due to the baking qualities of its flour and occupies a central strategic role in broadscale food security. Increasing yield and protein content in wheat grain is essential for safeguarding the human rights to sufficient and nutritious food. The quality and nutritional value of food products made from wheat flour largely depend upon the type and concentration of grain proteins [1,2]. As a consequence, total grain protein content (GPC) is one of the main determinants of both baking quality and the international market price of wheat [3]. The range of protein in wheat grains is usually

between 8% and 20%, accounting for less than 8–15% of the grain dry weight. Thousand kernel weight (TKW) is a measure of grain weight and, along with spike number per unit area and grain number per spike, is a main component of wheat total yield. Grain weight and, respectively, grain yield is formed mainly by the starch accumulation in the developing grain and, therefore, any starch gain in the endosperm if not accompanied by an adequate rise in nitrogen (N)/protein accumulation has a dilution effect that influences the concentration of grain protein and micronutrients [4]. Due to this important interaction, breeders frequently encounter conundrums when aiming for genotypes that combine high yield and high protein content in the grain [5–8].

Comparative studies on changes that have occurred in wheat varieties released or introduced since the middle of the 19th century showed increased grain yield and decreased protein over time [4,9]. This suggests that old germplasm (landraces and traditional varieties) can be screened for new genetic diversity and targeted for the construction of new varieties. Old germplasm grown in Bulgaria has some desirable traits, such as consistent yield, resistance to drought, high protein content, or good quality for making bread [10]. The semi-dwarf high-yielding varieties that emerged in the 1970s replaced the landraces and the tall varieties that were previously grown [11]. Most of this germplasm is now extinct, but some seed samples are preserved and reproduced in the major European seed gene banks [12].

Both GPC and TKW are quantitatively inherited and are controlled by multiple genes or quantitative trait loci (QTL) [13,14]. The impact of the environmental conditions on gene expression and the genotype-by-environment interactions further complicates the precise evaluation of these traits [15].

Genome-wide association studies (GWAS) detect genetic effects based on linkage disequilibrium (LD) in natural germplasm collections and have become effective tools for modern plant breeding [16]. Following the rapid development of DNA marker technologies, and in particular the advent of single nucleotide polymorphism (SNP) chips, association mapping has been used progressively to establish a strong connection between a genome-wide SNP and a trait of interest. Identifying marker-trait associations (MTAs) or quantitative trait nucleotides (QTNs) can make it easier for breeders to choose the best genotypes, reduce the breeding cycle, and achieve higher genetic gains. A number of recent GWAS studies reported genomic regions associated with GPC and yield components [14,17–31].

The association panels employed in these studies varied in terms of diversity level, genetic relatedness, and the nature of accessions. For instance, Kartseva et al. [27] used a diverse population of 255 accessions from 27 countries on five continents, and revealed novel stable genomic regions harboring GPC-associated markers on chromosomes 3A and 3B. QTL hotspots containing 165 significant MTAs for quality and agronomic traits were mapped on almost all chromosomes in an association panel of 170 diverse landraces from the Mediterranean region [31]. Using a set of 93 spring common wheat varieties and breeding lines adapted for cultivation in the Siberian region of the Russian Federation, another study reported eleven genomic regions associated with GPC, of which nine were physically mapped on chromosome 6A harboring the *NAM-A1* gene, homoeologous to the *Gpc-B1* (*NAM-B1*) gene [23]. Another recent investigation by Tyrka et al. [30] screened 168 Polish breeding lines of common winter wheat for a number of agronomic traits, and identified trait-associated markers for heading time, lodging resistance, plant height, and TKW. Therefore, it is essential to evaluate new large mapping populations with different underlying substructure, extents of genetic relatedness among individuals, and LD decays to identify novel QTNs or QTL for yield and quality traits. Additionally, these traits are influenced by environmental factors; therefore, detecting MTAs in multi-environment or multi-year studies is crucial for their application in marker-assisted breeding.

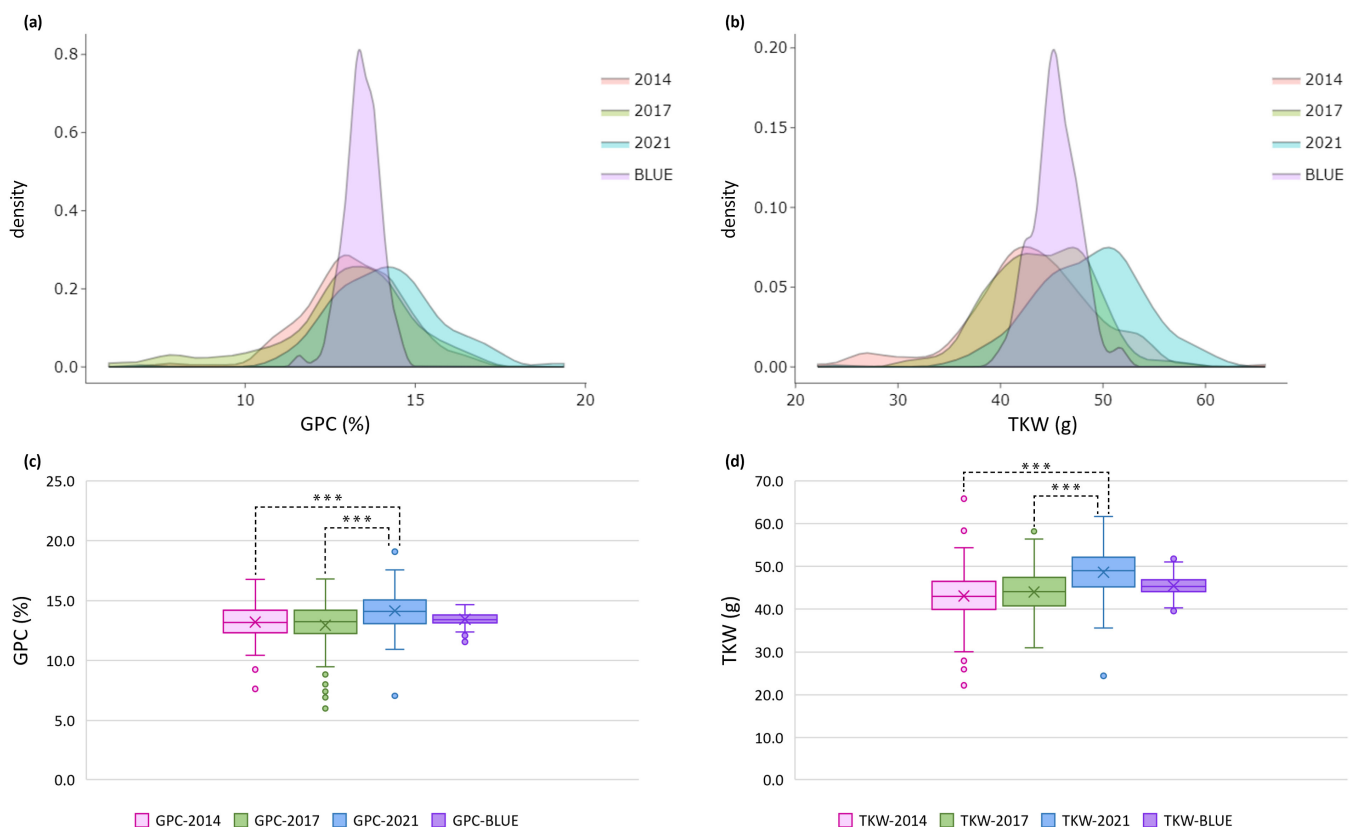
For this study, we assembled an association panel of advanced and historical varieties and local accessions from Bulgaria. Our fundamental focus was to explore the natural genetic variation for protein content in wheat grain and thousand kernel weight, and to

reveal the associated genetic determinants. We found a considerable amount of phenotypic variation for GPC and TKW with no apparent differences between modern high-productive varieties and old germplasm, nor did the traits significantly correlate with one another. Based on GWAS findings, we identified promising genomic regions for wheat improvement and uncovered possible candidate genes. These results provide information about the genetic resources available to breeders to improve grain yield and nutritional properties of wheat products, as well as an opportunity to develop closely associated markers to aid molecular breeding of new varieties.

## 2. Results

### 2.1. Phenotypic Variation

Descriptive statistics, frequency distribution, and boxplots showed wide phenotypic variation for both GPC and TKW over three growing seasons (harvests 2014, 2017, 2021) within the set of 179 wheat accessions with total average over the crop seasons of  $13.4 \pm 1.34\%$  for GPC, and  $45.2 \pm 4.28$  g for TKW (Tables 1, S1 and S2; Figure 1). The range of coefficients of variation (CV, %) across the individual years was similar for the two traits. Estimates of high broad-sense heritability ( $h^2$ ) for both traits showed moderate to high values in the individual growing seasons (from 0.64 to 0.78), and high values over the environments (0.82 for GPC and 0.81 for TKW). To exclude the impact of the growing season, we computed best linear unbiased estimator (BLUE) values for each accession, treating genotype as fixed and growing season as random effects. BLUEs varied across the years from 11.6 to 14.7%, on average 13.4% for GPC, and from 39.5 to 51.8 g, on average 45.4 g for TKW (Tables 1, S1 and S2; Figure 1).



**Figure 1.** Probability density and box plots for (a,c) grain protein content (GPC) and (b,d) thousand kernel weight (TKW) across three growing seasons and based on best linear unbiased estimators (BLUEs) in a set of 179 Bulgarian bread wheat accessions. \*\*\* is the significant differences between means at  $p < 0.001$ .

**Table 1.** Descriptive statistics for grain protein content (GPC) and thousand kernel weight (TKW) in a set of 179 Bulgarian bread wheat accessions evaluated over three growing seasons.

Trait	Env.	Mean *	Std. Dev.	Min.	Max.	CV %	$h^2$
GPC (%)	Sofia 2014	13.2 b	1.46	7.6	16.8	11.09	0.64
	Sofia 2017	12.9 b	2.06	6.0	16.8	15.90	0.78
	Sofia 2021	14.2 a	1.62	7.1	19.4	11.43	0.69
	Average	13.4	1.34	9.4	16.8	10.00	0.82
	BLUE	13.4	0.51	11.6	14.7	3.81	
TKW (g)	Sofia 2014	43.1 b	6.43	22.2	65.8	14.92	0.77
	Sofia 2017	44.0 b	4.72	31.0	58.2	10.73	0.64
	Sofia 2021	48.6 a	5.42	24.4	61.6	11.14	0.70
	Average	45.2	4.28	33.7	55.0	9.46	0.81
	BLUE	45.4	2.25	39.6	51.8	4.94	

\* different letters denote significant difference between the mean values at  $p < 0.001$ . Env. = environments; Std. Dev. = standard deviation; CV = coefficient of variation;  $h^2$  = broad-sense heritability; BLUE = best linear unbiased estimator.

The statistical analysis (Shapiro–Wilk test) indicates that the phenotypic data adheres to a normal distribution at a  $p$ -value threshold of 0.05. The ANOVA results explained the presence of broad phenotypic variation among genotypes for GPC and TKW, revealing highly significant effects of genotype, environment (growing season), and their interactions (Table 2). For both traits, no significant differences were noted only between 2014 and 2017 (Figure 1, Table S3).

**Table 2.** Factorial analysis of variance (ANOVA) for (a) grain protein content (GPC) and (b) thousand kernel weight (TKW) across three environments (growing seasons) for a set of 179 Bulgarian bread wheat accessions.

(a)						
Source of Variation	SS	df	MS	F	$p$ -Value	F Crit
Genotype (G)	962.685	178	5.408	3.010	0.0000	1.233
Environment (E)	145.483	2	72.741	40.485	0.0000	3.021
G $\times$ E	639.638	356	1.797	7.210	0.0000	3.320
Total	1747.805	536				
(b)						
Source of Variation	SS	df	MS	F	$p$ -Value	F Crit
Genotype (G)	9792.257	178	55.013	2.897	0.0000	1.233
Environment (E)	3128.650	2	1564.325	82.375	0.0000	3.021
G $\times$ E	6760.582	356	18.990	4.3643	0.0000	4.092
Total	19681.49	536				

SS = sum-of-squares; df = degrees of freedom; MS = mean squares; F = F-value; F crit = F-critical value.

To assess the trait consistency across the environments and to explain the relationships between GPC and TKW, Pearson’s correlation coefficient approach was used. Low to high positive Pearson’s correlation coefficients ( $r$ ) over the years were computed for GPC (spanning from 0.50 to 0.92) and TKW (ranging from 0.30 to 0.79) (Table 3). In general, consistency was noted across the growing seasons, with one exception—GPC in 2017 was not correlated with that in 2021. From the perspective of the data desirability for GWAS, correlation analysis was performed also with the BLUE mean values. The Pearson’s correlation coefficients ( $r$ ) computed based on BLUEs were positively significant ( $p < 0.001$ ) for both GPC and TKW. The two grain characteristics were not correlated across the growing seasons, as well as based on BLUEs (Table 3).

**Table 3.** Pearson’s correlation coefficients ( $r$ ) for grain protein content (GPC) and thousand kernel weight (TKW) among growing seasons (harvests in 2014, 2017, and 2021) and with the mean best linear unbiased estimator (BLUE) values in a set of 179 Bulgarian bread wheat accessions.

	GPC-2017	GPC-2021	GPC-BLUE	TKW-2014	TKW-2017	TKW-2021	TKW-BLUE
GPC-2014	0.69 ***	0.50 ***	0.93 ***	0.00	−0.09	−0.08	−0.05
GPC-2017		0.10	0.61 ***	0.04	0.01	−0.01	−0.02
GPC-2021			0.47 ***	0.16 *	−0.01	0.12	0.06
GPC-BLUE				−0.01	−0.10	−0.07	0.02
TKW-2014					0.39 ***	0.30 ***	0.60 ***
TKW-2017						0.53 ***	0.42 ***
TKW-2021							0.38 ***

\*, \*\*\* significant at  $p < 0.05$  and  $0.001$ , respectively.

According to a previous work, the studied population has a distinct structure based on the membership coefficient matrix (Q-matrix), calculated in STRUCTURE [32]. It is composed of three subpopulations (SPs), and only five genotypes are deemed to be admixed. The largest subpopulation—SP1 (109 entries) includes 6 old accessions and 103 modern varieties, most of them developed in the breeding centers in the Northern and Southern parts of the country. Of the 49 accessions in the SP2 subpopulation, 43 belong almost solely to the old germplasm, with the majority of them originating from Northern Bulgaria. The remaining 6 accessions are contemporary varieties. The smallest subpopulation—SP3—consists mostly of modern varieties (14), and 2 old accessions. Taking into account the distinct population structure, we tested the hypothesis of whether the old accessions differ significantly and consistently from the modern releases. Therefore, we compared the three SPs across the growing seasons. While the two subgroups containing modern varieties did not differ significantly across the years and with BLUEs concerning both GPC and TKW, the subgroup of old accessions (SP2) displayed variability across the environments. Subpopulation SP2 had significantly lower GPC compared to the SP3-varieties in 2021, but did not differ from both groups of modern releases in 2014 and 2017 and with BLUEs (Figure 2a). The group of old germplasm showed a higher TKW mean value in 2017 but a lower value in 2021 when compared to SP1, whereas the TKW mean BLUE of SP2 was significantly lower than that of SP3 (Figure 2b).

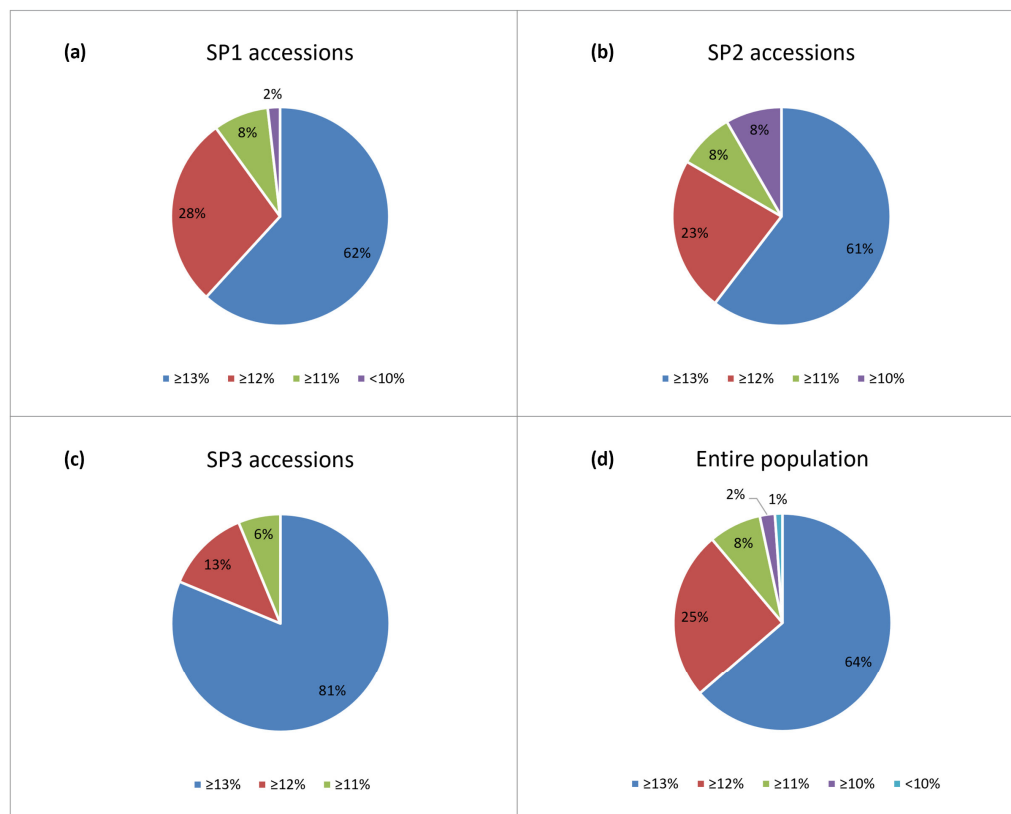
To distinguish genotypes that could be used for improving grain protein content in breeding programs, the accessions were classified according to [2] into five groups with protein  $\geq 13\%$  (Group 1),  $\geq 12\%$  (Group 2),  $\geq 11\%$  (Group 3), and  $>10\%$  (Group 4),  $<10\%$  (Group 5). The prevailing part of the accessions (90% and 94% of SP1 and SP3 modern releases, 84% of SP2 old accessions, or 89% of the entire population) fall into Groups 1 and 2 (Figure 3).

In order to evaluate the phenotypic variability of the population throughout the crop seasons, we initially determined the deviations from the respective yearly average GPC and TKW. Subsequently, these discrepancies were averaged for each accession throughout the three seasons. The influence of genotypes fluctuated per the environmental gradient, with the variance extending from 0.012 (2014) to 0.025 (2017) for GPC (Table S4), and from 0.011 (2017) to 0.022 (2014) for TKW (Table S5). The variance within the entire population over the years spanned from 0.0 to 0.108 for GPC, and from 0.0 to 0.101 for TKW.



**Figure 2.** Box plots for (a) grain protein content (GPC, in %) and (b) thousand kernel weight (TKW, in g) across three growing seasons and based on the best linear unbiased estimators (BLUEs) in the three subpopulations (SPs) distinguished by the Q-matrix [32] within a population of 179 Bulgarian bread wheat accessions. \*, \*\*\* are the significant differences between means at  $p < 0.05$  and 0.001, respectively.





**Figure 3.** Proportion of accessions (a–c) from the three subpopulations distinguished by the Q-matrix [32] and (d) within the entire population of 179 Bulgarian bread wheat accessions with respect to grain protein content (GPC); classification is according to [2].

## 2.2. Linkage Disequilibrium (LD) Estimation, Genotype Relatedness, and Significant Quantitative Trait Nucleotides (QTNs)

From the perspective of GWAS, LD was considered. This is essential to define the interval of highly associated SNPs and to identify the most significant loci [16]. Here, we calculated LD decay using the whole association panel irrespective of the genotype status (old vs. modern). The LD decay values varied from 1.5 to 3.0 Mbp on the individual chromosomes, with the highest value of LD decay in the D-genome (2.54 Mbp) and homoeologous group 3 (2.27 Mbp), on average 1.98 Mbp (Table S6).

High relatedness between the genotypes was demonstrated by the phylogenetic tree and the heat map of the values in the kinship matrix constructed from 17,083 SNPs (Figure S1).

In GWAS, controlling for population structure is a routine method that is especially relevant in this study because the genotypes comprise both historical and contemporary varieties as well as landraces. In an earlier study, the degree of diversity and population stratification were examined [32]. Three subpopulations were identified by STRUCTURE and the k-means clustering algorithm. The old germplasm was distinguished by both methods, as well as by principal component analysis. This work provides also a comprehensive comparative analysis of the old and modern accessions in terms of SNP-based genetic diversity and LD [32].

For both traits, the phenotypic data from three crop seasons and the genotypic data for the association panel that were already available from the 25K Infinium iSelect array (SGS Institut Fresenius GmbH TraitGenetics Section, Gatersleben, Germany) and described in [32] were used for association mapping. An examination of the variance showed substantial genotype-by-environment interactions (Table 2), despite the fact that both traits showed moderate to high broad-sense heritability estimates in each crop season (Table 1). Therefore, association mapping analysis was performed for each crop season, followed by

analysis based on BLUES using MLM (Q + K) model to account for population structure and kinship. The Q-Q plots showed deviations of the observed  $p$ -values from the null hypothesis suggesting there is a certain overestimation of the positive genomic signals (Figure S2). Therefore, in order to maintain a low false positive rate, the significance threshold was set at  $-\log_{10}(p\text{-value}) > \text{FDR}$  (false discovery rate),  $p < 0.01$ , and was calculated for each trait and environment.

GWAS for GPC evidenced 16 significant QTNs ( $-\log_{10} > \text{FDR}$ ,  $p < 0.01$ ) distributed on 11 chromosomes across the three environments and BLUES: 1B, 2A, 2B, 4B, 5A, 5B, 6A, 6B, 6D, 7A, and 7B (Table 4, Figure 4a). For the yield-related trait TKW, a total of 15 significant QTNs distributed on chromosomes 1A, 1B, 2A, 4A, 4D, 5B, 6A, and 6D were detected over the three crop seasons and based on BLUES (Table 5, Figure 4b). The proportion of phenotypic variance ( $R^2$ ) explained by the significantly associated markers was within the range 0.09 to 0.27 for GPC, and 0.12 to 0.23 for TKW; this range may have been upward biased to the relatively small population size. Seven QTNs presented positive additive effect on protein content spanning from 1.4% (*RFL\_Contig5739\_641*) to 8.1% (*AX-158545828*). Regarding TKW, four markers were identified as having positive additive effects on the trait, with *wsnp\_JD\_c4217\_5322858* having the most impact (8.4%).

**Table 4.** Quantitative trait nucleotides (QTNs) and candidate genes associated with grain protein content (GPC) detected in a population of 179 Bulgarian bread wheat accessions using the MLM (Q + K) model.

Marker	Chr	Position (bp)	$-\log_{10}(p)$	Effect	$R^2$	Gene ID	Annotation	Co-Located Loci <sup>a</sup>
GPC_2014								
<i>RAC875_rep_c98040_163</i>	1B	512,632,139	3.78	−1.054	0.084	<i>TraesCS1B01G294600</i>	GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase	
<i>BobWhite_c17047_268</i>	2B	11,077,394	3.85	−1.011	0.089	<i>TraesCS2B01G023800</i>	Ethylene receptor	
<i>wsnp_BE500291A_Ta_2_1</i>	5A	148,056,504	3.82	−0.889	0.087	<i>TraesCS5A01G101200</i>	1-acyl-sn-glycerol-3-phosphate acyltransferase	
<i>Tdurum_contig54917_597</i>	6B	32,324,229	3.73	1.357	0.083	<i>TraesCS6B01G052900</i>	Kelch repeat-containing protein	[33]
GPC_2017								
<i>CAP11_c601_120</i>	4B	38,791,353	4.21	−2.456	0.096	<i>TraesCS4B01G050200</i>	Histidine kinase	
<i>AX-95148678</i>	4B	39,015,957	4.15	−2.446	0.095	<i>TraesCS4B01G050400</i>	UDP-glucose 6-dehydrogenase	
<i>AX-94513493</i>	4B	39,045,208	4.21	−2.456	0.096	<i>TraesCS4B01G050500</i>	UV-B-induced protein, chloroplastic	
<i>BS00109912_51</i>	7A	653,925	4.36	2.874	0.101	NA		
GPC_2021								
<i>RFL_Contig5739_641</i>	5B	531,538,634	4.11	1.412	0.092	<i>TraesCS5B01G350900</i>	Plant regulator RWP-RK family protein, putative	[34]
<i>IACX2946</i>	6A	599,046,570	4.16	−2.172	0.092	<i>TraesCS6A01G378000</i>	Protein kinase family protein	[23,33]
<i>Tdurum_contig46670_911</i>	6A	599,050,921	4.16	2.172	0.092	<i>TraesCS6A01G378100</i>	Diphthine-ammonia ligase	[23,33]
<i>AX-94649878</i>	6D	453,347,731	4.86	−1.954	0.111	<i>TraesCS6D01G362900</i>	NAC domain protein	
GPC_BLUE								
<i>Ra_c22880_760</i>	2A	498,105,752	8.38	7.918	0.217	<i>TraesCS2A01G289800</i>	Alpha-amylase	
<i>AX-158545828</i>	2A	501,850,514	8.46	8.069	0.219	<i>TraesCS2A01G291200</i>	Mitochondrial import inner membrane translocase subunit TIM22	
<i>Tdurum_contig59467_433</i>	7A	715,692,069	9.94	−9.349	0.267	NA		
<i>AX-94519723</i>	7B	121,890,360	8.57	7.372	0.222	<i>TraesCS7B01G105700</i>	DNA-binding storekeeper protein-related transcriptional regulator	

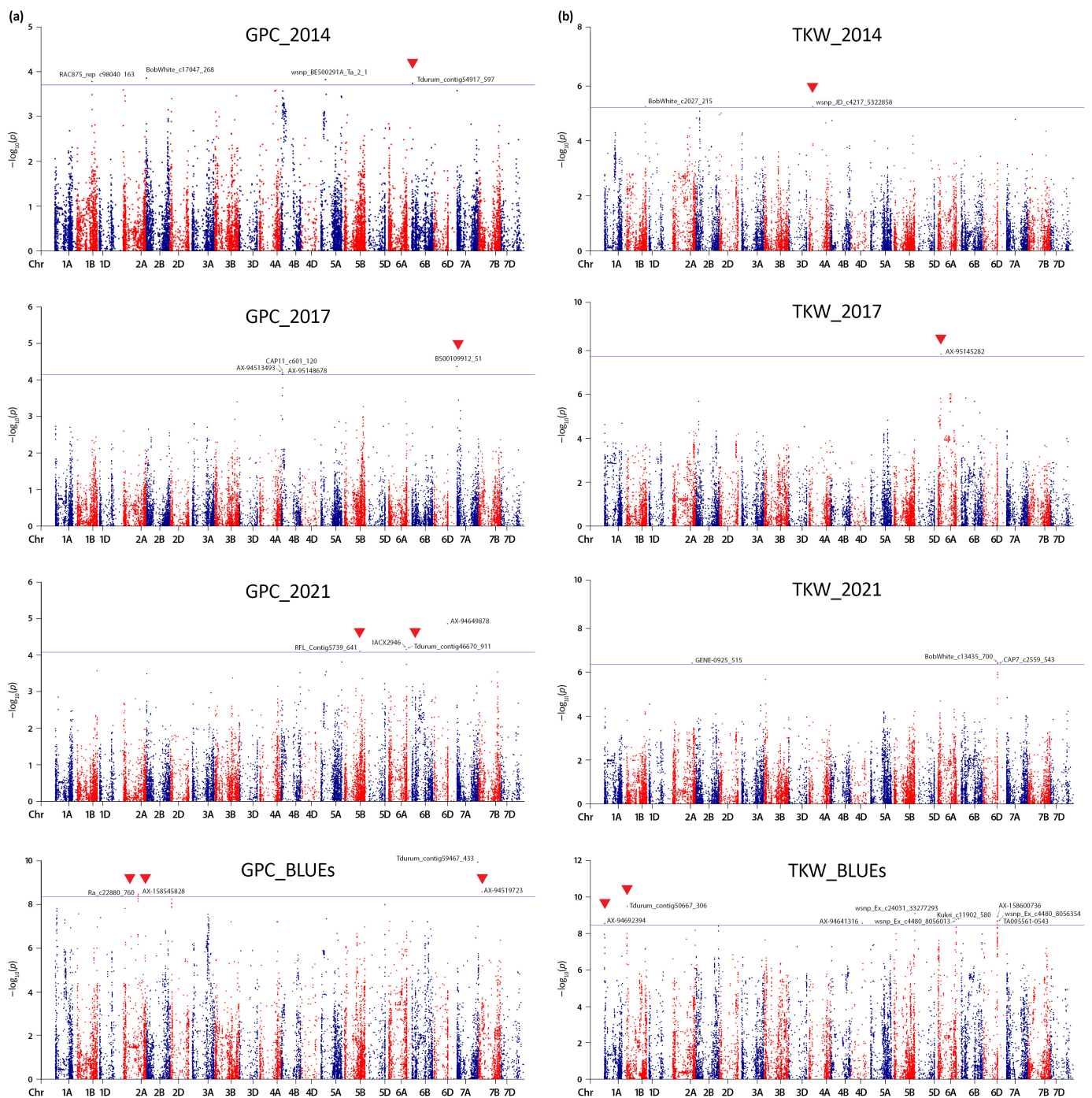
<sup>a</sup> Previously reported significant marker/QTL in close proximity to the QTNs identified in the current study.

No environmentally stable QTNs were detected for GPC and TKW, which confirms that both grain traits are significantly influenced by the crop seasons (environments) and demonstrate the presence of genotype-by-environment interactions.

In our analysis, we examined the significant markers associated with GPC and TKW across the three distinct subpopulations, as differentiated by the Q-matrix in a previous study [32]. Our observations indicate that among the 16 loci significantly associated with GPC, only locus *BS00109912\_51* exclusively (100%) carries allele C in the old accessions (SP2 subpopulation). However, the other allele (T) is still present in the old germplasm, albeit at a lower frequency (20% of all accessions with this allele) (Table S7). In the case of 8 out of 15 significant loci associated with TKW, one allele is predominantly (95–100%) present



in the SP1 and SP3 subpopulations, which are mainly composed of modern varieties. The other allele is also highly represented among the modern releases (Table S8).



**Figure 4.** Manhattan plots of the genome-wide association scan for (a) grain protein content (GPC) and (b) thousand kernel weight (TKW) based on the MLM (Q + K) model and the phenotypic data from three growing seasons, along with the calculated BLUE values in a set of 179 Bulgarian bread wheat accessions. The blue color line depicts the threshold of  $-\log_{10}(p) = \text{FDR}$ , determined for each trait and environment. The significant quantitative trait nucleotides (QTNs) are above the blue line. QTNs with positive additive effects on the traits are marked with red color triangles.

**Table 5.** Quantitative trait nucleotides (QTNs) and candidate genes associated with thousand kernel weight (TKW) detected in a population of 179 Bulgarian bread wheat accessions using the MLM (Q + K) model.

Marker	Chr	Position (bp)	$-\log_{10}(p)$	Effect	R <sup>2</sup>	Gene ID	Annotation	Co-Located Loci <sup>a</sup>
TKW_2014								
<i>BobWhite_c2027_215</i>	1B	640,555,504	5.27	−10.268	0.124	<i>TraesCS1B01G415400</i>	TSA: Wollemia nobilis Ref_Wollemi_Transcript_14910_3291 transcribed RNA sequence	[21],[35]
<i>wsnp_JD_c4217_5322858</i>	4A	140,291,399	5.25	8.388	0.123	<i>TraesCS4A01G115300</i>	Small nuclear ribonucleoprotein	[36]
TKW_2017								
<i>AX-95145282</i>	6A	100,766,008	7.84	4.780	0.199	NA		
TKW_2021								
<i>GENE-0925_515</i>	2A	656,820,630	6.42	−4.526	0.143	NA		
<i>BobWhite_c13435_700</i>	6D	471,008,534	6.43	−4.534	0.144	NA		[37]
<i>CAP7_c2559_543</i>	6D	471,017,962	6.40	−4.519	0.143	<i>TraesCS6D01G402800</i>	Zinc finger CCCH domain-containing protein 19	[37]
TKW_BLUEs								
<i>AX-94692394</i>	1A	7,186,246	8.54	1.927	0.209	<i>TraesCS1A01G012600</i>	Leucine-rich repeat receptor-like protein kinase family protein	[21],[35]
<i>Tdurum_contig50667_306</i>	1B	20,588,032	9.49	2.705	0.238	<i>TraesCS1B01G041200</i>	F-box protein-like	
<i>AX-94641316</i>	4D	374,832,658	8.55	−2.485	0.210	<i>TraesCS4D01G219400</i>	Regulator of Vps4 activity in the MVB pathway protein, putative PGR5-like protein 1A, chloroplastic	[35],[37]
<i>wsnp_Ex_c24031_33277293</i>	5B	707,131,836	9.10	−3.002	0.226	<i>TraesCS5B01G560400</i>	Lysine-specific histone demethylase 1-3-like protein	[35],[37],[38]
<i>wsnp_Ex_c4480_8056013</i>	6A	611,487,998	8.76	−2.084	0.216	<i>TraesCS6A01G405500</i>	Lysine-specific histone demethylase 1-3-like protein	[35],[37],[38]
<i>Kukri_c11902_580</i>	6A	611,491,408	8.76	−2.085	0.216	<i>TraesCS6A01G405500</i>	Lysine-specific histone demethylase 1-3-like protein	[35],[37],[38]
<i>wsnp_Ex_c4480_8056354</i>	6D	464,946,552	8.67	−2.078	0.214	<i>TraesCS6D01G389200</i>	Lysine-specific histone demethylase 1-3-like protein	[37]
<i>AX-158600736</i>	6D	464,948,739	8.70	−2.078	0.215	<i>TraesCS6D01G389200</i>	Lysine-specific histone demethylase 1-3-like protein	[37]
<i>TA005561-0543</i>	6D	465,050,368	8.48	−2.065	0.211	NA		[37]

<sup>a</sup> Previously reported significant markers/QTL coinciding with (underlined references) or in close proximity to the QTN identified in the current study.

### 2.3. Potential Candidate Genes

In order to hypothesize potential candidate genes that underlie the analyzed traits, we examined the 16 significant QTNs detected for GPC and the 15 significant QTNs detected for TKW. The search for candidate genes was performed based on the physical position of the prominent QTNs on *Triticum aestivum* L. cv. Chinese Spring reference genome extended by the LD interval estimated for each chromosome based on LD decay. Within the defined regions, associated with GPC and TKW, we retrieved a total of 493 and 639 high-confidence annotated genes, respectively (Tables S9 and S10). These genes were evaluated as potential candidate genes. The genes directly hit by the significant SNPs, along with their putative functions, are presented in Tables 4 and 5.

In the following section, we look at the genes with significant QTNs of positive additive effects, as well as at several pertinent candidates detected in the LD-defined intervals on both sides of the QTNs. Relevant potential genes linked to GPC and TKW encode storage proteins, proteins implicated in senescence- and germination-associated proteolysis, carbohydrate synthesis, protein synthesis and trafficking, and auxin biosynthesis. Products of certain candidate genes are transcriptional and posttranscriptional regulators, or transmembrane factors that can sense, transduce, and transmit signals.

## 3. Discussion

### 3.1. Phenotypic Variation

To meet the growing demand for sufficient amounts and quality of food and overcome the challenges posed by environmental changes, it is essential to effectively utilize the available genetic resources of bread wheat [39]. Here, we used a collection of old and modern bread wheat accessions, and explored the genetic variation of TKW and GPC, the two most important characters determining yield and end-use quality, and eventually, the eco-

onomic value of bread wheat. The study revealed large phenotypic variability for the target traits with highly significant contributions of genotype, environment, and genotype-by-environment interactions. The observed moderate to high values of broad-sense heritability and the consistency of trait records evidenced by the significant correlations between the crop seasons suggest that a considerable part of the variation is due to inherent genetic differences among the accessions. These results agree with similar findings in bread and durum wheat for GPC [19,25,27,40], and TKW [13,41]. Given that the panel contains historical and contemporary varieties released or collected between 1925 and 2010, it is possible that the recorded genetic variance—especially for TKW—is inflated. This is because throughout the course of nearly nine decades, genetic factors have contributed significantly to increases in yield and yield-related attributes. The landmark of these genetic gains for the Bulgarian wheat collection is the introduction of semi-dwarfing genes in the 1970s [11]. The pleiotropic effects that these genes exert on plant responsiveness to N applications, on photosynthetic rates, and on the accumulation of carbohydrates in the grain ultimately have improved grain yield.

This study shows that a large proportion (89%) of the accessions have protein above 12%, and are suitable for preparing leavened breads [2]. The average grain protein values (13.4%) are similar to or higher than the reported information on protein contents in other bread wheat collections [19,27,28,31,34]. Additional research revealed that some accession sets had greater protein values (up to 20%), demonstrating the impact of genotype-by-environment interactions [22,23,33].

Wheat yield and quality are affected by climatic factors (temperature, precipitation, drought type) [42]. In contrast to the abundant rainfall that was recorded at the experimental site from April to July in 2014, stretching from heading time to maturity, the amount of rain that fell in May and June 2017 was consistently less than the average for the climate (Figure S3), which suggests a sustained moderate drought during anthesis and grain filling. Prolonged water insufficiency early in grain development reduces the number of amyloplasts and endosperm cells, which lowers the capacity of starch accumulation, and ultimately lowers grain weight [43,44]. In 2021, a modest drought during anthesis and a more severe final drought were seen (Figure S3). Minor postanthesis water stress can speed up grain filling in wheat by boosting the activity of key catalytic enzymes that convert sucrose to starch, and by remobilizing nonstructural carbohydrates from the vegetative tissues to the grain [44]. The effects of these patterns of drought (timing and severity) explain well the significantly higher mean TKW value obtained in 2021 than those in 2017. The considerably lower mean TKW recorded in 2014 is consistent with research showing that prolonged soil wetness after anthesis restricts the amount of assimilates available to growing grains, hence diminishing the development of grain yield [45]. The observed variance in the protein content in the collection under study may potentially be explained by the precipitation oscillations among the three crop seasons. In our study, we found a significant difference ( $p < 0.001$ , Table 1) in the average protein values for 2014 (13.2%) and 2021 (14.2%), when prolonged waterlogging or terminal drought periods, respectively, were suggested during late vegetation. A decline in GPC has been associated with waterlogging [46]. Conversely, during a drought, an increase in the seed proteins has been shown [47], possibly connected to modified carbon (C) partitioning and, hence, to a shift in the C/N ratio, that favors greater N-assimilation [48].

Significant genotype-by-environment interactions were seen (Table 2), and the genotypic effects varied along with the environmental gradient (Tables S4 and S5). It is interesting to note, that throughout the growing seasons varieties Mustang, Bozhana, and Levent displayed consistently high positive deviation from the average GPC and high values of TKW. These accessions are potential sources for concurrent improvement of the two traits. Ancestral history indicates that the Mustang variety is descended from the old accession Yubilejna-3, which also exhibits consistently high protein levels in the grain. Additionally, its pedigree involves a hybrid derived from *Agropyron* sp., a wheatgrass that has been shown to contribute to seed storage proteins in wheat-*Agropyron* introgressions [49,50].

The lack of association between TKW and GPC in our data using Pearson's approach shows that there is little to no decrease in seed weight in the presence of the identified loci for GPC. Similarly, no significant correlations were found between GPC and TKW across environments in studies of wheat lines derived from wild emmer [51,52]. This observation is consistent with findings by Oury and Godin [6] that genotype-by-environment interactions for grain yield and GPC may obscure the strong genetic background of the yield–protein interrelationship. Thus, Lindeque et al. [53] detected limited significant correlations between grain yield and protein content in a study of wheat accessions of various yielding capacities grown in environments of different precipitation trends. This result holds promise for the simultaneous genetic enhancement of the two traits.

Historical germplasm, such as landraces and traditional varieties, are an important source for bringing new genes into contemporary crops [31,54,55]. Therefore, it is strategically important to characterize these genetic resources in order to properly utilize them in prebreeding. Our comparative analysis of variation for GPC showed that although the old germplasm (SP2) was characterized in general by high values of grain protein concentrations (Figure 3, Table S1), these values were lower than those of modern releases or did not differ notably from them (Figure 2). Moreover, the percentage of high-protein genotypes ( $\geq 12\%$ ) was higher in the two modern subpopulations, SP1 (90%) and SP3 (94%), compared to the old accessions of SP2 (84%). The idea that the older germplasm had superior yield-related metrics than the more recent varieties was also not supported by the observed variability in mean TKW values of SP2-accessions across the growing seasons. One likely reason is that a lot of contemporary Bulgarian varieties are either descended from old accessions, or have highly productive and high-protein Russian and Serbian ancestors in their pedigrees as Table S1 and the research by Kartseva et al. [27] show. These results suggest that wheat breeders have managed to construct improved varieties in terms of the studied traits.

### 3.2. Genomic Regions Associated with Grain Protein Content (GPC) and Thousand Kernel Weight (TKW)

Significant verified SNPs or genomic regions linked to GPC and TKW have been detected on all wheat chromosomes, according to recent research [14,35,56,57]. We compared the strongly associated SNPs found in this study to previously published loci or markers using the IWGSC RefSeq v.1.0 map as a reference.

Our analysis confirmed some known trait-associated genomic regions, or found associated markers in proximity (distance 0.1–6.5 Mbp) to previously described loci [21,23,33–38] (Tables 4 and 5). For example, the GPC-associated marker *Tdurum\_contig46670\_911* on chromosome 6A (position 599,050,921 bp) has been reported as an environmentally specific one in a study on a Siberian wheat collection [23]. The same marker is also located inside a known QTL (q6A-3), defined by the interval 599.2–602.9 Mbp [33], and stays two Mbp apart from locus *QGpc.ipg-6A.4* reported by Leonova et al. [23] (Table 4). Similarly, six TKW-associated QTNs detected on chromosomes 1B, 5B, 6A, and 6D overlapped with previously reported QTL [21,35,37] (Table 5).

Certain significant QTNs have not been previously linked to the studied traits, suggesting that these loci are novel. New genomic loci that seem promising for grain protein improvement in wheat harbor markers *Ra\_c22880\_760* and *AX-158545828* on chromosome 2A, *AX-94519723* on chromosome 7B, and *BS00109912\_51* on chromosome 7A, presenting protein increasing effects up to 8.1%. Two potentially novel loci with positive additive effects on TKW are defined in this study. The linked markers are *Tdurum\_contig50667\_306* on chromosome 1B, and *AX-95145282* on chromosome 6A. Some of the trait-associated loci were detected with BLUEs, whereas the rest are environmentally specific and need further validation.

The comparative analysis of the historical and contemporary genotypes, focusing on the significant markers associated with GPC and TKW, revealed little to no divergence of the old germplasm. This observation is consistent with molecular findings from previous



research on Bulgarian wheat germplasm which utilized microsatellites [58] and SNPs [32]. These studies demonstrated that wheat breeders have maintained the high levels of genetic diversity found in the older germplasm since the 1970s. Furthermore, a shift in some prevailing alleles from the older accessions to the modern ones seems to have occurred, with some unique or rare alleles disappearing and others being introduced.

### 3.3. Putative Candidate Genes Related to Grain Protein Content (GPC) and Thousand Kernel Weight (TKW)

Grain production and nutrient content of cereal crops are governed by the buildup of nutrients in the grain near the end of the plant lifespan [59]. The quantity and quality of these nutrients is greatly affected by the remobilization of C and N from vegetative tissues to developing grain during senescence, as well as by the subsequent accumulation of resources for seed germination. All these processes are accompanied by expressional changes in a vast number of genes that ultimately impact grain output and protein content [60]. We searched for potential candidate genes at suggestively novel loci, as well as at loci that verified previously established associations. The next section covers potential genes in genomic regions harboring QTNs with positive additive effects on the traits.

Our findings highlighted seven QTNs with positive effects on grain protein, of which the two QTNs with the highest effect (ca. 8% each) are located in the interval 498,105,752–501,850,514 bp on chromosome 2A. Marker *Ra\_c22880\_760* resides within *TraesCS2A01G289800*, a gene translated into alpha-amylase. Alpha-amylase is one of the primary enzymes responsible for starch degradation to fuel early germinating grain [61]. The gene *TraesCS2A01G291200* hit by marker *AX-158545828* is annotated as mitochondrial import inner membrane translocase subunit TIM22. This protein is a PRAT (preprotein and amino acid transporters) family member and an important constituent of the mitochondrial protein import machinery [62]. Although the effect of this gene has not previously been documented in wheat, the functions of the *Arabidopsis* orthologous gene during seed development have been recently reported [63].

The gene *TraesCS7B01G105700* on chromosome 7B hit by marker *AX-94519723* encodes for a DNA-binding storekeeper protein-related transcriptional regulator (STKR). The STKR was the first B-box motif plant-specific binding protein identified as a putative regulator of expression of patatin, the most abundant storage protein of potato tubers [64]. In transgenic *Arabidopsis* plants, overexpression of the orthologous protein (STKR1) was shown to result in reduced growth, a delay in flowering, attenuated senescence, exhausted carbohydrate pool, and amino acids accumulation [65] and, therefore, is related to nutrient source-to-sink remobilization.

One QTN (*BS00109912\_51*) is located on chromosome 7A, possibly in a noncoding interval. Within the LD-expanded genomic region to the left and right of the significant SNP, we retrieved two genes encoding cysteine-proteases. These enzymes are among the most abundant proteases activated during leaf senescence that are essential for N remobilization from senesced leaves to developing seeds [66].

We found prospective putative candidate genes for GPC improvement in the genomic region on chromosome 6A harboring marker *Tdurum\_contig46670\_911* (position 599,050,921 bp). The gene *TraesCS6A01G378100* translated into Diphthine-ammonia ligase has not been previously associated with effects in wheat. This enzyme is evolutionarily conserved in eukaryotes and catalyzes the last step in the conversion of an L-histidine residue in the translation elongation factor eEF2 to diphthamide [67]. Interestingly, there is a gene close to *TraesCS6A01G378100*, encoding translation elongation factor EF1A. The elongation factors (EFs) are key mediators in the peptide chains elongation during protein synthesis [68]. Nearby *TraesCS6A01G378100* we found two genes encoding NAC-domain-containing proteins. NAC genes, especially *NAM-B1* (*Gpc-B1*) together with its homoeologs *NAM-A1* and *NAM-D1* have been shown to regulate the transcriptional changes during leaf senescence in wheat, affecting N and nutrient source-to-sink remobilization [69,70]. Marker *Tdurum\_contig46670\_911* identified in the current study is situated 0.1 Mb away from a GPC

locus (*QGpc.icg-6A.4*), described in the distal part of the long arm of chromosome 6A [23]. Since *NAM-A1* gene is located on the short arm of 6A [23], the *NAC* genes identified here close to *TraesCS6A01G378100* are, therefore, distinct.

The QTN *RFL\_Contig5739\_641* showing increase on GPC of ca. 1.4% resides inside the gene *TraesCS5B01G350900* coding for plant regulator RWP-RK family protein. RWP-RKs represent a small family of transcription factors that are unique to plants. The NLPs (NIN-like proteins) subfamily of RWP-RKs have been confirmed as master regulators of the nitrate signaling and the expression of a series of genes for nitrate transport and assimilation, including nitrate-transporter and nitrate reductase genes [71,72]. Interesting putative candidates for grain protein are four genes detected nearby *TraesCS5B01G350900* that were translated into flavin-containing monooxygenase, and one gene coding for subtilisin-like protease. In plants, the flavin-containing monooxygenase (FMO) gene superfamily is implicated in the auxin biosynthesis [73]. Auxin plays a pivotal role as a regulator, modulating a variety of cellular processes that are integral to seed formation. Genes coding for auxin-related FMO had steadily increasing expression levels during the wheat grain expanding phase, indicating a potential role in grain development [74]. Subtilisin-like proteases contribute to the degradation of reserve proteins during germination [75]. Genes for subtilisin-like proteases have been proposed as possible candidates for GPC [19] and seed longevity [76] in wheat.

Another QTN, which contributes to approximately a 1.4% increase in protein, is located within the gene *TraesCS6B01G052900* on chromosome 6B. This gene codes for a Kelch repeat-containing protein, a subclass of F-box proteins. F-box proteins are important players in the ubiquitin/26S proteasome system [77]. Protein breakdown, one of the most significant catabolic processes associated with leaf senescence, plays a key role in nutrient recycling, especially N. Consequently, it is closely linked to grain filling in cereals [75]. Chen et al. [78] reported on an F-box protein containing a Kelch repeat motif (*OsFBK12*) that regulates leaf senescence, seed size, and grain number in rice (*Oryza sativa*). The expression of wheat *TaKFB* genes encoding Kelch repeat F-box proteins was shown to increase during the pigmentation stage of grain development [79]. Close to *TraesCS6B01G052900* we found another gene coding for subtilisin-like protease, discussed above.

Four QTNs were identified that had positive effects on TKW. The QTN on chromosome 4A with the highest effect (ca. 8%) co-locates with *TraesCS4A01G115300*, a gene encoding small nuclear ribonucleoprotein (snRNP). The snRNPs, along with over 300 other splicing factors, form the spliceosome, the enzyme that catalyzes the splicing of the primary transcripts to mature mRNAs, an essential step in gene expression [80]. Lopato et al. [81] showed that the wheat gene *TaRSZ38*, a homologue of the *Arabidopsis* *AtRSZ33* splicing factor, regulates posttranscriptional processes in wheat grain with an expression peak in the immature grain. A relevant candidate gene for TKW, located 682,883 bp apart from *TraesCS4A01G115300* is translated into aminotransferase. A genome-wide analysis in bread wheat has identified 15 Tryptophan Aminotransferase of *Arabidopsis*1/Tryptophan Aminotransferase-Related (*TAA1/TAR*) genes involved in auxin biosynthesis. These findings highlight the potential of the gene *TaTAR2.1-3A* for genetically enhancing grain yield. It has been demonstrated that overexpression of this gene leads to increased plant height, spike number, grain yield, and aerial N accumulation under various N supplies [82].

Our analysis did not detect a candidate gene directly associated with QTN AX-95145282 (position: 100,766,008 bp, chromosome 6A) having ca. 5% increasing effect on TKW. The gene search within the LD-expanded interval found 16 genes encoding leucine-rich repeat receptor-like protein kinase family protein (LRR-RLKs) and a squamosa promoter-binding protein-like (SBP domain) transcription factor family protein. LRR-RLKs are the largest group of receptor-like kinases in plants implicated in plant growth, development, and stress responses [83]. The effects of this protein family on yield-related traits in wheat have not been reported, but overexpression of the rice gene *LRK1* has been shown to increase cell proliferation and several yield components, resulting in a 27% increase in total grain yield per plant [84]. Studies in rice [85] and wheat [86] reported on putative candidate



genes for grain size and TKW, encoding for squamosa promoter binding protein-like (SPL), a star player for plant growth and development [87]. Interestingly, the *Arabidopsis AtSPL9* gene was reported as a potential regulator of nitrate transporters and nitrate reductase [72] thereby possibly regulating the N assimilation.

The QTN *Tdurum\_contig50667\_306*, located on chromosome 1B (position: 20,588,032 bp; ca. 3% trait increase) is inside the gene *TraesCS1B01G041200*, that translates into an F-box protein-like. The role of this super protein family in plant cells was briefly covered previously in this section. Recent studies suggested that the wheat gene *WAP0-A1*, also known as *TaAPO-A1*, encoding an F-box protein is associated with variation for grain yield and quality traits [8,88,89]. In close proximity to *TraesCS1B01G041200*, we found another gene that seems to play a significant role in explaining the variation for TKW. This gene encodes a developmentally regulated G protein. Heterotrimeric G proteins in plants are membrane-attached signal transducers known to control diverse cellular processes, such as cell proliferation, hormonal responses, growth, and development [90]. G protein-mediated regulation of grain size has been reported in *Arabidopsis* [91], rice [90,92], barley [93], and could therefore impact the harvestable yield in crop plants. Moreover, the plant G protein complex was shown to regulate N signaling and N use efficiency in rice [94].

The QTN *AX-94692394* on chromosome 1A with the lowest increasing effect on TKW (ca. 2%) resides within gene *TraesCS1A01G012600*, translated into leucine-rich repeat receptor-like protein kinase family protein. Earlier in this section, we provided short information about the possible functions of LRR-RLKs. Interestingly, we discovered a putative candidate gene, *TraesCS1A01G010900*, in the LD interval close to *TraesCS1A01G012600* annotated as gliadin/LMW glutenin/bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain. Similarly, Giancaspro et al. [86] and Schierenbeck et al. [95] identified candidate genes coding for the primary prolamins in TKW-coding regions. This finding is not surprising given that gliadins and glutenins make up the majority of endosperm reserve proteins in wheat grain. Others also reported on such function of suggestive candidate genes underlying marker associations with grain architecture traits and TKW [36,95]. Another relevant candidate in the 1A-genomic region (*TraesCS1A01G010500*) encodes cellulose synthase-like protein (Csl). This enzyme controls the cellulose biosynthesis thereby contributing to the cell wall organization. A recent genome-wide analysis of the *Csl* gene family in bread wheat revealed 108 *Csl* genes, and some of them were expressed in the grain [96]. Interestingly, a candidate gene for the wheat tiller inhibition gene (*tin*) that reduces tillering and increases grain weight was predicted to encode a Csl protein [97].

## 4. Materials and Methods

### 4.1. Plant Material

The plant material consisted of 179 winter wheat (*Triticum aestivum* L.) accessions that originated from Bulgaria, of which 129 were modern varieties, and 50 were old accessions (historic varieties of tall stature and landraces). Seeds from the old accessions were procured from the seed gene banks at the Leibniz Institute for Plant Genetics and Crop Research (IPK), Gatersleben, Germany, and the Crop Research Institute, Prague, Czech Republic. According to the available information, the period of early breeding and research expeditions to aggregate seeds from landraces and historic varieties for gene bank collections expands from 1925 to 1970. Seeds from the contemporary varieties released until 2010 were made available from the two major breeding centers in Bulgaria (Dobrudzha Agricultural Institute, General Toshevo and the Institute of Plant Genetic Resources, Sadovo, Bulgaria), as well as by breeders. Information about botanical variety, status (modern vs. old), year of release, and known genealogy is given in Table S1.

### 4.2. Phenotyping

The seed material was collected from field experiments conducted in Sofia, Bulgaria (42°41' N, 23°19' E) during three crop seasons (2013/14, 2016/17, and 2020/21), denoted by the year of harvest (2014, 2017, and 2021). For each growing season, the accessions were

sown in a random design in double 1 m-long rows and two replications. The soil type at the experimental field is leached vertisol, pH 6.1, with 3.1% humus, 1420 mg total N per kg soil, of which 18 mg inorganic N. Plants received 120 kg N/ha as ammonium nitrate in two split doses, 40 kg/ha two weeks after sowing and 80 kg/ha before stem extension growth stage. For pest control, standard agricultural practices were applied. The average monthly temperature ranged between  $-5.8^{\circ}\text{C}$  and  $23.2^{\circ}\text{C}$ , during the vegetation cycles (October–July). The monthly precipitation ranged from 4.7 mm to 150.5 mm, with average annual precipitation of 682 mm, which is around 20% higher than the climate norm for Sofia. Information on the monthly weather statistics for the region of the experimental field is presented in Figure S3.

Plant material was hand-harvested and hand-threshed. For each genotype and crop season, the thousand kernel weight (TKW) was determined based on grain number and grain mass of 10 main spikes per replication. The grain protein content (GPC) was measured in three pooled whole-grain powder samples (1.0 g each) for each accession and year. Before analyses, seeds were dried to a constant weight and ground to a fine powder with IKA Tube Mill Control (IKA Werke GmbH & Co., Staufen, Germany). The total N concentration was determined using a UDK 159 Automatic Kjeldahl Nitrogen Protein Analyzer (Velp Scientifica, Usmate, MB, Italy), and the corresponding GPC values (in %) were automatically inferred from the N values, according to [98].

#### 4.3. Statistical Analyses

The normal distribution of the empirical data was checked by the Shapiro–Wilk test using the online software MVApp (<https://mvapp.kaust.edu.sa>) (accessed on 12 January 2024). The significant differences in GPC and TKW among accessions, growing seasons, and interaction effects between genotype and environment were tested using analysis of variance (ANOVA). The best linear unbiased estimators (BLUEs) for each accession across the growing seasons were obtained to eliminate the environmental impact by assuming the genotype as a fixed effect and the growing season as a random effect. The relationships for GPC and TKW among the growing seasons and with the BLUEs of these traits were tested by Pearson's correlation coefficients ( $r$ ). The correlation coefficients  $r$  were calculated from the empirical phenotypic data obtained in each growing season and from the mean BLUE values.

Broad sense heritability  $h^2$  for GPC and TKW was calculated using the formula

$$h^2 = \frac{\sigma_G^2}{\left(\sigma_G^2 + \frac{\sigma_E^2}{nE}\right)}$$

where  $\sigma_G^2$  is the genotype variance,  $\sigma_E^2$  is the variance of the residual, and  $nE$  is the number of environments (growing seasons). Heritability in each environment was estimated using the same formula, where  $nE$  in the denominator is the number of replications in a given environment.

All phenotypic data analyses were accomplished using STATISTICA 14 [99].

#### 4.4. Association Mapping and Candidate Gene Search

Before performing marker-trait association analysis, the population stratification, the genetic relatedness among population entries, and the LD were considered. The genotypic data for the association panel of 179 accessions were already available from the 25K Infinium iSelect array (SGS Institut Fresenius GmbH TraitGenetics Section, Gatersleben, Germany) and described in [32]. Population structure was modeled using the Bayesian clustering algorithm in STRUCTURE 2.3.4, and the membership coefficients (Q-values) were determined [32]. Kinship was estimated as a similarity matrix (K) from 17,083 SNPs based on the method developed by VanRaden [100] using the Genomic Association and

Prediction Integrated Tool (GAPIT) in R package. The values of LD decay (in Mbp) were determined for each chromosome as described in [32].

Considering the distinctive population structure and the high genetic relatedness among population individuals, GWAS was undertaken with a mixed linear model (MLM) Q + K model to control pseudo associations [101]. Hence, the filtered set of 17,083 SNPs, phenotypic data from the three crop seasons, and the calculated BLUE mean values for the traits, along with population structure (Q values) and kinship similarity matrix (K) as covariates were used for association mapping analysis in TASSEL v.5.

To analyze our data, we applied another powerful algorithm—the fixed and random model circulating probability unification (FarmCPU)—through GAPIT in R. The FarmCPU model uses the advantages of the MLM and stepwise regression (fixed effect model) and, at the same time, overcomes their disadvantages, which makes this approach very accurate for GWAS by effectively controlling the rate of false positive signals [102]. The FarmCPU algorithm delivered altogether three significant associations for the two traits, while the MLM output involved a large number of genomic signals. To test how the two models fit the data, we examined the quantile–quantile (Q-Q) plots, which show the distributions of important  $p$ -values (expected vs. observed  $-\log_{10}(p\text{-values})$ ). Based on the number of significant SNPs (QTNs) and the Q-Q plots, we present here the results obtained by the MLM (Q + K) model. In order to reduce the number of false positive signals, we set the threshold of statistically significant QTNs at  $-\log_{10}(p\text{-value}) > \text{FDR}$  (false discovery rate), at  $p < 0.01$ . FDR was calculated for each trait and environment.

To draw the Manhattan plots and Q-Q plots, R package qqman was employed [103].

Additive effects and  $R^2$  (percent phenotypic variation) of QTNs were estimated in TASSEL v.5 and Excel 16 (Microsoft). The percentage of phenotypic variation explained by each QTN ( $R^2$ ) was calculated as the difference of  $R^2$  with and without the strongest associated SNP.

For the significant trait-associated SNPs, flanking regions within the LD-estimated interval on either side of the marker were searched for candidate genes. Within these regions, high-confidence putative candidate genes were predicted by blasting against the cv. Chinese Spring reference genome IWGSC RefSeq v1.0. [104]. Gene annotations, gene ontologies (GOs), InterPros, and details for the potential candidate genes were obtained using EnsemblPlants, and Persephone web-based platforms ([http://plants.ensembl.org/Triticum\\_aestivum/Info/Index](http://plants.ensembl.org/Triticum_aestivum/Info/Index); <https://web.persephonesoft.com/?data>) (accessed on 12 March 2024).

## 5. Conclusions

The GWAS revealed significant QTNs underlying both studied traits. Among these, seven GPC-associated loci as well as four TKW-related loci had notably positive effect on the traits, thereby representing promising genomic regions for use in breeding projects aiming at improving grain protein and yield, respectively. While some of the loci confirmed already published ones, others might be new. In addition, to our knowledge, the targeted GPC-linked regions on chromosomes 2A, 7A, and 7B, as well as two TKW-associated intervals on chromosomes 1B and 6A, have not been previously linked to the studied traits. Based on their gene functional annotation, a few of the many high-confidence genes that were retrieved from the trait-associated genomic regions were selected as intriguing candidates. Annotations to senescence- and germination-related proteolysis, peptide chain elongation and protein translocation, synthesis of storage proteins, structural carbohydrates, and auxin, as well as a number of transcriptional regulatory factors and signal transducers, are present in these genes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants13081084/s1>, Figure S1: Heat map of kinship matrix with dendrogram based on 17,083 SNPs; Figure S2: Q-Q plots of the genome-wide association scan for (a) GPC and (b) TKW; Figure S3: Climate characteristics at the experimental site, Sofia, Bulgaria (42°41' N, 23°19' E) during three crop seasons (2013/14, 2016/17 and 2020/21). The precipitation is

related to the climate norm (1961–1990) accepted by the Bulgarian National Institute of Meteorology and Hydrology, Ministry of Environment and Water; Table S1: Information about 179 bread wheat accessions from Bulgaria and summary statistics for GPC from three crop seasons (harvests 2014, 2017, 2021), average over the seasons, and the average BLUE values; Table S2: Information about 179 bread wheat accessions from Bulgaria, and summary statistics for TKW from three crop seasons (harvests 2014, 2017, 2021), average over the seasons, and the average BLUE values; Table S3: Analysis of variance of the single factor “Growing season” on GPC and TKW in a set of 179 bread wheat accessions from Bulgaria in three growing seasons (harvests 2014, 2017, 2021); Table S4: Variability and stability of GPC in a set of 179 bread wheat accessions from Bulgaria; Table S5: Variability and stability of TKW in a set of 179 bread wheat accessions from Bulgaria; Table S6: Values of linkage disequilibrium (LD) decay for each chromosome, homoeologous group and genome in a set of 179 bread wheat accessions from Bulgaria; Table S7: Comparative analysis at the significant GPC-associated SNP loci; Table S8: Comparative analysis at the significant TKW-associated SNP loci; Table S9: Candidate genes for GPC; Table S10: Candidate genes for TKW.

**Author Contributions:** Conceptualization, S.M. and T.K.; formal analysis, V.A. and A.M.A.; investigation, T.K., K.K. and D.D.; resources, T.K., K.P., S.M. and A.B.; data curation, T.K. and V.A.; writing—original draft preparation, S.M.; writing—review and editing, A.M.A., M.A.R.A., A.B. and K.K.; visualization, V.A., T.K., S.M. and M.A.R.A.; supervision, S.M. and A.M.A.; project administration, S.M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Bulgarian National Science Fund, grant number KP-06-N31/17.

**Data Availability Statement:** Data are contained within the article and its Supplementary Materials.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Veraverbeke, W.S.; Delcour, J.A. Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. *Crit. Rev. Food Sci. Nutr.* **2002**, *42*, 179–208. [CrossRef] [PubMed]
2. Peña, R.J. Wheat for Bread and Other Foods. In *Bread Wheat. Improvement and Production*; FAO Plant Production and Protection Series No. 30; Curtis, B.S., Rajaram, S., Gómez Macpherson, H., Eds.; Food and Agriculture Organization of the United Nations: Rome, Italy, 2002. Available online: <https://www.fao.org/3/y4011e/y4011e0w.htm#bm32> (accessed on 10 December 2023).
3. Barneix, A.J. Physiology and biochemistry of source-regulated protein accumulation in the wheat grain. *J. Plant Physiol.* **2007**, *164*, 581–590. [CrossRef]
4. Shewry, P.R.; Hassall, K.L.; Grausgruber, H.; Andersson, A.A.M.; Lampi, A.-M.; Piironen, V.; Rakszegi, M.; Ward, J.L.; Lovegrove, A. Do modern types of wheat have lower quality for human health? *Nutr. Bull.* **2020**, *45*, 362–373. [CrossRef]
5. Cooper, M.; Woodruff, D.R.; Philips, I.G.; Basford, K.E.; Gilmour, A.R. Genotype-by-management interactions for grain yield and grain protein concentration of wheat. *Field Crops Res.* **2001**, *70*, 87–88. [CrossRef]
6. Oury, F.X.; Godin, C. Yield and grain protein concentration in bread wheat: How to use the negative relationship between the two characters to identify favourable genotypes? *Euphytica* **2007**, *157*, 45–57. [CrossRef]
7. Laidig, F.; Piepho, H.P.; Rentel, D.; Drobek, T.; Meyer, U.; Huesken, A. Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in german official variety trials and on-farm during 1983–2014. *Theor. Appl. Genet.* **2017**, *130*, 223–245. [CrossRef] [PubMed]
8. Geyer, M.; Mohler, V.; Hartl, L. Genetics of the inverse relationship between grain yield and grain protein content in common wheat. *Plants* **2022**, *11*, 2146. [CrossRef]
9. Poudel, R.; Bhinderwala, F.; Morton, M.; Powers, R.; Rose, D.J. Metabolic profiling of historical and modern wheat cultivars using proton nuclear magnetic resonance spectroscopy. *Sci. Rep.* **2021**, *11*, 3080. [CrossRef]
10. Majdrakov, P. For wheat from Pavlikeni. *Seed Prod.* **1945**, *4*, 132–141. (In Bulgarian)
11. Ganeva, G.; Korzun, V.; Landjeva, S.; Tsenov, N.; Atanasova, M. Identification, distribution and effects on agronomic traits of the semi-dwarfing Rht alleles in Bulgarian bread wheat cultivars. *Euphytica* **2005**, *145*, 305–315. [CrossRef]
12. Börner, A.; Landjeva, S.; Nagel, M.; Rehman Arif, M.A.; Allam, M.; Agacka, M.; Doroszewska, T.; Lohwasser, U. Plant genetic resources for food and agriculture (PGRFA)—Maintenance and research. *Genet. Plant Physiol.* **2014**, *4*, 13–21.
13. Garcia, M.; Eckermann, P.; Haefele, S.; Satija, S.; Sznajder, B.; Timmins, A.; Baumann, U.; Wolters, P.; Mather, D.E.; Fleury, D. Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS ONE* **2019**, *14*, e0211730. [CrossRef]
14. Paina, C.; Gregersen, P.L. Recent advances in the genetics underlying wheat grain protein content and grain protein deviation in hexaploid wheat. *Plant Biol.* **2023**, *25*, 661–670. [CrossRef]



15. Robert, N.; Hennequet, C.; Berard, P. Dry matter and nitrogen accumulation in wheat kernel: Genetic variation in rate and duration of grain filling. *J. Genet. Breed.* **2001**, *55*, 297–306.
16. Alqudah, A.M.; Sallam, A.; Baenziger, P.S.; Börner, A. GWAS: Fast-forwarding gene identification and characterization in temperate cereals: Lessons from barley—A review. *J. Adv. Res.* **2020**, *22*, 119–135. [\[CrossRef\]](#)
17. Daba, S.D.; Tyagi, P.; Brown-Guedira, G.; Mohammadi, M. Genome-wide association studies to identify loci and candidate genes controlling kernel weight and length in a historical United States wheat population. *Front. Plant Sci.* **2018**, *9*, 1045. [\[CrossRef\]](#)
18. Muhammad, A.; Hu, W.; Li, Z.; Li, J.; Xie, G.; Wang, J.; Wang, L. Appraising the genetic architecture of kernel traits in hexaploid wheat using GWAS. *Int. J. Mol. Sci.* **2020**, *21*, 5649. [\[CrossRef\]](#)
19. Muqaddasi, Q.H.; Brassac, J.; Ebmeyer, E.; Kollers, S.; Korzun, V.; Argillier, O.; Stiewe, G.; Plieske, J.; Ganal, M.W.; Röder, M.S. Prospects of GWAS and predictive breeding for European winter wheat's grain protein content, grain starch content, and grain hardness. *Sci. Rep.* **2020**, *10*, 12541. [\[CrossRef\]](#)
20. Alemu, A.; El Baouchi, A.; El Hanafi, S.; Kehel, Z.; Eddakhir, K.; Tadesse, W. Genetic analysis of grain protein content and dough quality traits in elite spring bread wheat (*Triticum aestivum*) lines through association study. *J. Cereal Sci.* **2021**, *100*, 103214. [\[CrossRef\]](#)
21. Rufo, R.; López, A.; Lopes, M.S.; Bellvert, J.; Soriano, J.M. Identification of quantitative trait loci hotspots affecting agronomic traits and high-throughput vegetation indices in rainfed wheat. *Front. Plant Sci.* **2021**, *12*, 735192. [\[CrossRef\]](#)
22. Suliman, S.; Alemu, A.; Abdelmula, A.A.; Badawi, G.H.; Al-Abdallat, A.; Tadesse, W. Genome-wide association analysis uncovers stable QTLs for yield and quality traits of spring bread wheat (*Triticum aestivum*) across contrasting environments. *Plant Gene* **2021**, *25*, 100269. [\[CrossRef\]](#)
23. Leonova, I.N.; Kiseleva, A.A.; Berezhnaya, A.A.; Stasyuk, A.I.; Likhenko, I.E.; Salina, E.A. Identification of QTLs for grain protein content in Russian spring wheat varieties. *Plants* **2022**, *11*, 437. [\[CrossRef\]](#)
24. White, J.; Sharma, R.; Balding, D.; Cockram, J.; Mackay, I.J. Genome-wide association mapping of Hagberg falling number, protein content, test weight, and grain yield in U.K. wheat. *Crop Sci.* **2022**, *62*, 965–981. [\[CrossRef\]](#)
25. Alomari, D.Z.; Schierenbeck, M.; Alqudah, A.M.; Alqahtani, M.D.; Wagner, S.; Rolletschek, H.; Borisjuk, L.; Röder, M.S. Wheat grains as a sustainable source of protein for health. *Nutrients* **2023**, *15*, 4398. [\[CrossRef\]](#)
26. Guo, Y.; Wang, G.; Guo, X.; Chi, S.; Yu, H.; Jin, K.; Huang, H.; Wang, D.; Wu, C.; Tian, J.; et al. Genetic dissection of protein and starch during wheat grain development using QTL mapping and GWAS. *Front. Plant Sci.* **2023**, *14*, 1189887. [\[CrossRef\]](#)
27. Kartseva, T.; Alqudah, A.M.; Aleksandrov, V.; Alomari, D.Z.; Doneva, D.; Arif, M.A.R.; Börner, A.; Misheva, S. Nutritional genomic approach for improving grain protein content in wheat. *Foods* **2023**, *12*, 1399. [\[CrossRef\]](#)
28. Krishnappa, G.; Khan, H.; Krishna, H.; Devate, N.B.; Kumar, S.; Mishra, C.N.; Parkash, O.; Kumar, S.; Kumar, M.; Mamrutha, H.M.; et al. Genome-wide association study for grain protein, thousand kernel weight, and normalized difference vegetation index in bread wheat (*Triticum aestivum* L.). *Genes* **2023**, *14*, 637. [\[CrossRef\]](#)
29. Sharma, A.; Arif, M.A.R.; Shamshad, M.; Rawale, K.S.; Brar, A.; Burgueño, J.; Shokat, S.; Ravinder Kaur, R.; Vikram, P.; Srivastava, P.; et al. Preliminary dissection of grain yield and related traits at differential nitrogen levels in diverse pre-breeding wheat germplasm through association mapping. *Mol. Biotechnol.* **2023**, *65*, 116–130. [\[CrossRef\]](#)
30. Tyrka, M.; Krajewski, P.; Bednarek, P.T.; Rączka, K.; Drzazga, T.; Matysik, P.; Martofel, R.; Woźna-Pawlak, U.; Jasińska, D.; Niewińska, M.; et al. Genome-wide association mapping in elite winter wheat breeding for yield improvement. *J. Appl. Genet.* **2023**, *64*, 377–391. [\[CrossRef\]](#)
31. Yannam, V.R.R.; Lopes, M.; Guzman, C.; Soriano, J.M. Uncovering the genetic basis for quality traits in the Mediterranean old wheat germplasm and phenotypic and genomic prediction assessment by cross-validation test. *Front. Plant Sci.* **2023**, *4*, 1127357. [\[CrossRef\]](#)
32. Aleksandrov, V.; Kartseva, T.; Alqudah, A.M.; Kocheva, K.; Tasheva, K.; Börner, A.; Misheva, S. Genetic diversity, linkage disequilibrium and population structure of Bulgarian bread wheat assessed by genome-wide distributed SNP markers: From old germplasm to semi-dwarf cultivars. *Plants* **2021**, *10*, 1116. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Yang, Y.; Chai, Y.; Zhang, X.; Lu, S.; Zhao, Z.; Wei, D.; Chen, L.; Hu, Y.-G. Multi-locus GWAS of quality traits in bread wheat: Mining more candidate genes and possible regulatory Network. *Front. Plant Sci.* **2020**, *11*, 1091. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Rathan, N.D.; Krishna, H.; Ellur, R.K.; Sehgal, D.; Govindan, V.; Ahlawat, A.K.; Krishnappa, G.; Jaiswal, J.P.; Singh, J.B.; Saiprasad, S.V.; et al. Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Sci. Rep.* **2022**, *12*, 7037. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Cao, S.; Xu, D.; Hanif, M.; Xia, X.; He, Z. Genetic architecture underpinning yield component traits in wheat. *Theor. Appl. Genet.* **2020**, *33*, 1811–1823. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Eltaher, S.; Sallam, A.; Emara, H.A.; Nower, A.A.; Salem, K.F.M.; Börner, A.; Baenziger, P.S.; Mourad, A.M.I. Genome-wide association mapping revealed SNP alleles associated with spike traits in wheat. *Agronomy* **2022**, *12*, 1469. [\[CrossRef\]](#)
37. Yang, Y.; Amo, A.; Wei, D.; Chai, Y.; Zheng, J.; Qiao, P.; Cul, C.; Lu, S.; Chen, L.; Hu, Y. Large-scale integration of meta-QTL and genome-wide association study discovers the genomic regions and candidate genes for yield and yield-related traits in bread wheat. *Theor. Appl. Genet.* **2021**, *134*, 3083–3109. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Chidzanga, C.; Daniel Mullan, D.; Roy, S.; Baumann, U.; Garcia, M. Nested association mapping-based GWAS for grain yield and related traits in wheat grown under diverse Australian environments. *Theor. Appl. Genet.* **2022**, *135*, 4437–4456. [\[CrossRef\]](#) [\[PubMed\]](#)

39. Arif, M.A.R.; Waheed, M.Q.; Lohwasser, U.; Shokat, S.; Alquddah, A.M.; Volkmar, C.; Börner, A. Genetic insight into the insect resistance in bread wheat exploiting the untapped natural diversity. *Front. Genet.* **2022**, *13*, 898905. [\[CrossRef\]](#)
40. Rathana, N.D.; Krishnappa, G.; Singh, A.-M.; Govindan, V. Mapping QTL for phenological and grain-related traits in a mapping population derived from high-zinc-biofortified wheat. *Plants* **2023**, *12*, 220. [\[CrossRef\]](#)
41. Mulugeta, B.; Tesfaye, K.; Ortiz, R.; Johansson, E.; Hailesilassie, T.; Hammenhag, C.; Hailu, F.; Geleta, M. Marker-trait association analyses revealed major novel QTLs for grain yield and related traits in durum wheat. *Front. Plant Sci.* **2023**, *13*, 1009244. [\[CrossRef\]](#)
42. Akram, S.; Arif, M.A.R.; Hameed, A. A GBS-based GWAS analysis of adaptability and yield traits in bread wheat (*Triticum aestivum* L.). *J. Appl. Genet.* **2020**, *62*, 27–41. [\[CrossRef\]](#)
43. Wan, C.; Dang, P.; Gao, L.; Wang, J.; Tao, J.; Qin, X.; Feng, B.; Gao, J. How does the environment affect wheat yield and protein content response to drought? A meta-analysis. *Front. Plant Sci.* **2022**, *13*, 896985. [\[CrossRef\]](#)
44. Teng, Z.; Chen, Y.; Meng, S.; Duan, M.; Ye, N.; Zhang, J. Environmental stimuli: A major challenge during grain filling in cereals. *Int. J. Mol. Sci.* **2023**, *24*, 2255. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Ma, S.; Hou, J.; Wang, Y.; Wang, M.; Zhang, W.; Fan, Y.; Huang, Z. Post-flowering soil waterlogging curtails grain yield formation by restricting assimilates supplies to developing grains. *Front. Plant Sci.* **2022**, *13*, 944308. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Herzog, M.; Striker, G.G.; Colmer, T.D.; Pedersen, O. Mechanisms of waterlogging tolerance in wheat—A review of root and shoot physiology. *Plant Cell Environ.* **2016**, *39*, 1068–1086. [\[CrossRef\]](#)
47. Gooding, M.J.; Ellis, R.H.; Shewry, P.R.; Schofield, J.D. Effects of restricted water availability and increased temperature on the grain filling, drying and quality of winter wheat. *J. Cereal Sci.* **2003**, *37*, 295–309. [\[CrossRef\]](#)
48. Sehgal, D.; Mondal, S.; Guzman, C.; Garcia Barrios, G.; Franco, C.; Singh, R.; Dreisigacker, S. Validation of candidate gene-based markers and identification of novel loci for thousand-grain weight in spring bread wheat. *Front. Plant Sci.* **2019**, *10*, 1189. [\[CrossRef\]](#)
49. Han, F.; Liu, B.; Fedak, G.; Liu, Z. Genomic constitution and variation in five partial amphiploids of wheat–*Thinopyrum intermedium* as revealed by GISH, multicolor GISH and seed storage protein analysis. *Theor. Appl. Genet.* **2004**, *109*, 1070–1076. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Georgieva, M.; Seps, A.; Tyankova, N.; Molnár-Láng, M. Molecular cytogenetic characterization of two high protein wheat–*Thinopyrum intermedium* partial amphiploids. *J. Appl. Gen.* **2011**, *52*, 269–277. [\[CrossRef\]](#)
51. Fatiukha, A.; Filler, N.; Lupo, I.; Lidzbarsky, G.; Klymiuk, V.; Korol, A.B.; Pozniak, C.; Fahima, T.; Krugman, T. Grain protein content and thousand kernel weight QTLs identified in a durum × wild emmer wheat mapping population tested in five environments. *Theor. Appl. Genet.* **2020**, *133*, 119–131. [\[CrossRef\]](#)
52. Liu, J.; Huang, L.; Wang, C.; Liu, Y.; Yan, Z.; Wang, Z.; Xiang, L.; Zhong, X.; Gong, F.; Zheng, Y.; et al. Genome-wide association study reveals novel genomic regions associated with high grain protein content in wheat lines derived from wild emmer wheat. *Front. Plant Sci.* **2019**, *10*, 464. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Lindeque, R.C.; van Biljon, A.; Labuschagne, M.T. Defining associations between grain yield and protein quantity and quality in wheat from the three primary production regions of South Africa. *J. Cereal Sci.* **2018**, *79*, 294e302. [\[CrossRef\]](#)
54. Cseh, A.; Pocza, P.; Kiss, T.; Balla, K.; Berki, Z.; Horváth, A.; Kuti, C.; Karsai, I. Exploring the legacy of Central European historical winter wheat landraces. *Sci. Rep.* **2021**, *11*, 23915. [\[CrossRef\]](#) [\[PubMed\]](#)
55. López-Fernández, M.; Pascual, L.; Faci, I.; Fernández, M.; Ruiz, M.; Benavente, E.; Giraldo, P. Exploring the end-use quality potential of a collection of spanish bread wheat landraces. *Plants* **2021**, *10*, 620. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Liu, H.; Mullan, D.; Zhang, C.; Zhao, S.; Li, X.; Zhang, A.; Lu, Z.; Wang, Y.; Yan, G. Major genomic regions responsible for wheat yield and its components as revealed by meta-QTL and genotype–phenotype association analyses. *Planta* **2020**, *252*, 65. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Saini, P.; Sheikh, I.; Saini, D.K.; Mir, R.R.; Dhaliwal, H.S.; Tyagi, V. Consensus genomic regions associated with grain protein content in hexaploid and tetraploid wheat. *Front. Genet.* **2022**, *13*, 1021180. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Landjeva, S.; Korzun, V.; Ganeva, G. Evaluation of genetic diversity among Bulgarian winter wheat (*Triticum aestivum* L.) varieties during the period 1925–2003 using microsatellites. *Genet. Res. Crop Evol.* **2006**, *53*, 1605–1614. [\[CrossRef\]](#)
59. Borrill, P.; Harrington, S.A.; Simmonds, J.; Uauy, C. Identification of transcription factors regulating senescence in wheat through gene regulatory network modelling. *Plant Physiol.* **2019**, *180*, 1740–1755. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Sultana, N.; Islam, S.; Juhasz, A.; Ma, W. Wheat leaf senescence and its regulatory gene network. *Crop J.* **2021**, *9*, 703–717. [\[CrossRef\]](#)
61. Whan, A.; Dielen, A.-S.; Mieog, J.; Bowerman, A.F.; Robinson, H.M.; Byrne, K.; Colgrave, M.; Larkin, P.J.; Howitt, C.A.; Morell, M.K.; et al. Engineering α-amylase levels in wheat grain suggests a highly sophisticated level of carbohydrate regulation during development. *J. Exp. Bot.* **2014**, *65*, 5443–5457. [\[CrossRef\]](#)
62. Kolli, R.; Soll, J.; Carrie, C. Plant mitochondrial inner membrane protein insertion. *Int. J. Mol. Sci.* **2018**, *19*, 641. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Zhang, Y.; Hu, Y.; Wang, Z.; Lin, X.; Li, Z.; Ren, Y.; Zhao, J. The translocase of the inner mitochondrial membrane 22-2 is required for mitochondrial membrane function during *Arabidopsis* seed development. *J. Exp. Bot.* **2023**, *74*, 4427–4448. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Zourelidou, M.; De Torres-Zabala, M.; Smith, C.; Bevan, M.W. Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of patatin expression. *Plant J.* **2002**, *30*, 489–497. [\[CrossRef\]](#) [\[PubMed\]](#)



65. Nietzsche, M.; Guerra, T.; Alseekh, S.; Wiermer, M.; Sonnewald, S.; Fernie, A.R.; Börnke, F. STOREKEEPER RELATED1/G-element binding protein (STKR1) interacts with protein kinase SnRK1. *Plant Physiol.* **2018**, *176*, 1773–1792. [\[CrossRef\]](#)
66. Hajibarat, Z.; Saidi, A. Senescence-associated proteins and nitrogen remobilization in grain filling under drought stress condition. *J. Genet. Eng. Biotechnol.* **2022**, *20*, 101. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Su, X.; Lin, Z.; Chen, W.; Jiang, H.; Zhang, S.; Lin, H. Chemogenomic approach identified yeast YLR143W as diphthamide synthetase. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19983–19987. [\[CrossRef\]](#)
68. Xu, L.; Zhang, L.; Liu, Y.; Sod, B.; Li, M.; Yang, T.; Gao, T.; Yang, Q.; Long, R. Overexpression of the elongation factor MteF1A1 promotes salt stress tolerance in *Arabidopsis thaliana* and *Medicago truncatula*. *BMC Plant Biol.* **2023**, *23*, 138. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Uauy, C.; Distelfeld, A.; Fahima, T.; Blechl, A.; Dubcovsky, J. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* **2006**, *314*, 1298–1301. [\[CrossRef\]](#)
70. Andleeb, T.; Borrill, P. Wheat NAM genes regulate the majority of early monocarpic senescence transcriptional changes including nitrogen remobilisation genes. *G3 Genes Genomes Genet.* **2023**, *13*, jkac275. [\[CrossRef\]](#)
71. Konishi, M.; Yanagisawa, S. *Arabidopsis* NIN-like transcription factors have a central role in nitrate signalling. *Nat. Commun.* **2013**, *4*, 1617. [\[CrossRef\]](#)
72. Mahmoud, D.; Pandey, R.; Sathee, L.; Dalal, M.; Singh, M.P.; Chinnusamy, V. Regulation of expression of genes associated with nitrate response by osmotic stress and combined osmotic and nitrogen deficiency stress in bread wheat (*Triticum aestivum* L.). *Plant Physiol. Rep.* **2020**, *25*, 200–215. [\[CrossRef\]](#)
73. Schlaich, N.L. Flavin-containing monooxygenases in plants: Looking beyond detox. *Trends Plant Sci.* **2007**, *12*, 412–418. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Guan, J.; Wang, Z.; Liu, S.; Kong, X.; Wang, F.; Sun, G.; Geng, S.; Mao, L.; Zhou, P.; Li, A. Transcriptome analysis of developing wheat grains at rapid expanding phase reveals dynamic gene expression patterns. *Biology* **2022**, *11*, 281. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Diaz-Mendoza, M.; Diaz, I.; Martinez, M. Insights on the proteases involved in barley and wheat grain germination. *Int. J. Mol. Sci.* **2019**, *20*, 2087. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Arif, M.A.R.; Börner, A. An SNP based GWAS analysis of seed longevity in wheat. *Cereal Res. Comm.* **2020**, *48*, 149–156. [\[CrossRef\]](#)
77. Vierstra, R.D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 385–397. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Chen, Y.; Xu, Y.; Luo, W.; Li, W.; Chen, N.; Zhang, D.; Chong, K. The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. *Plant Physiol.* **2013**, *163*, 1673–1685. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Hong, M.J.; Kim, D.Y.; Choi, H.I.; Seo, Y.W.; Kim, J.B. Isolation and characterization of kelch repeat-containing F-box proteins from colored wheat. *Mol. Biol. Rep.* **2020**, *47*, 1129–1141. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Patel, S.B.; Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* **2008**, *36*, 6482–6493. [\[CrossRef\]](#)
81. Lopato, S.; Borisjuk, L.; Milligan, A.S.; Shirley, N.; Bazanova, N.; Langridge, P. Systematic identification of factors involved in post-transcriptional processes in wheat grain. *Plant Mol. Biol.* **2006**, *62*, 637–653. [\[CrossRef\]](#)
82. Shao, A.; Ma, W.; Zhao, X.; Hu, M.; He, X.; Teng, W.; Li, H.; Tong, Y. The auxin biosynthetic TRYPTOPHAN AMINOTRANSFERASE RELATED TaTAR2. 1-3A increases grain yield of wheat. *Plant Physiol.* **2017**, *174*, 2274–2288. [\[CrossRef\]](#)
83. Dufayard, J.F.; Bettembourg, M.; Fischer, I.; Droc, G.; Guiderdoni, E.; Périn, C.; Chantret, N.; Diévar, A. New insights on leucine-rich repeats receptor-like kinase orthologous relationships in angiosperms. *Front. Plant Sci.* **2017**, *8*, 243806. [\[CrossRef\]](#)
84. Zha, X.; Luo, X.; Qian, X.; He, G.; Yang, M.; Li, Y.; Yang, J. Over-expression of the rice LRK1 gene improves quantitative yield components. *Plant Biotechnol. J.* **2009**, *7*, 611–620. [\[CrossRef\]](#)
85. Si, L.; Chen, J.; Huang, X.; Gong, H.; Luo, J.; Hou, Q.; Zhou, T.; Lu, T.; Zhu, J.; Shangguan, Y.; et al. OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* **2016**, *48*, 447–456. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Giancaspro, A.; Giove, S.L.; Zacheo, S.A.; Blanco, A.; Gadaleta, A. Genetic variation for protein content and yield-related traits in a durum population derived from an inter-specific cross between hexaploid and tetraploid wheat cultivars. *Front. Plant Sci.* **2019**, *10*, 1509. [\[CrossRef\]](#)
87. Chen, X.; Zhang, Z.; Liu, D.; Zhang, K.; Li, A.; Mao, L. SQUAMOSA promoter-binding protein-like transcription factors: Star players for plant growth and development. *J. Integr. Plant Biol.* **2010**, *52*, 946–951. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Kuzay, S.; Xu, Y.; Zhang, J.; Katz, A.; Pearce, S.; Su, Z.; Fraser, M.; Anderson, J.A.; Brown-Guedira, G.; DeWitt, N.; et al. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theor. Appl. Genet.* **2019**, *132*, 2689–2705. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Voss-Fels, K.P.; Keeble-Gagnère, G.; Hickey, L.T.; Tibbits, J.; Nagorny, S.; Hayden, M.J.; Wittkop, B. High-resolution mapping of rachis nodes per rachis, a critical determinant of grain yield components in wheat. *Theor. Appl. Genet.* **2019**, *132*, 2707–2719. [\[CrossRef\]](#)
90. Xu, R.; Li, N.; Li, Y. Control of grain size by G protein signaling in rice. *J. Integr. Plant Biol.* **2019**, *61*, 533–540. [\[CrossRef\]](#)
91. Li, S.; Liu, Y.; Zheng, L.; Chen, L.; Li, N.; Corke, F.; Lu, Y.; Fu, X.; Zhu, Z.; Bevan, M.W.; et al. The plant-specific G protein  $\gamma$  subunit AGG3 influences organ size and shape in *Arabidopsis thaliana*. *New Phytol.* **2012**, *194*, 690–703. [\[CrossRef\]](#)
92. Sun, S.; Wang, L.; Mao, H.; Shao, L.; Li, X.; Xiao, J.; Ouyang, Y.; Zhang, Q. A G-protein pathway determines grain size in rice. *Nat. Commun.* **2018**, *9*, 851. [\[CrossRef\]](#)

93. Wendt, T.; Holme, I.; Dockter, C.; Preuß, A.; Thomas, W.; Druka, A.; Waugh, R.; Hansson, M.; Braumann, I. HvDep1 is a positive regulator of culm elongation and grain size in barley and impacts yield in an environment-dependent manner. *PLoS ONE* **2016**, *11*, e0168924. [[CrossRef](#)]
94. Sun, H.; Qian, Q.; Wu, K.; Luo, J.; Wang, S.; Zhang, C.; Ma, Y.; Liu, Q.; Huang, X.; Yuan, Q.; et al. Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat. Gen.* **2014**, *46*, 652–656. [[CrossRef](#)] [[PubMed](#)]
95. Schierenbeck, M.; Alqudah, A.M.; Lohwasser, U.; Tarawneh, R.A.; Simón, M.R.; Börner, A. Genetic dissection of grain architecture related traits in a winter wheat population. *BMC Plant Biol.* **2021**, *21*, 417. [[CrossRef](#)]
96. Kaur, S.; Dhugga, K.S.; Beech, R.; Singh, J. Genome-wide analysis of the *cellulose synthase-like (Csl)* gene family in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2017**, *17*, 193. [[CrossRef](#)]
97. Hyles, J.; Vautrin, S.; Pettolino, F.; MacMillan, C.; Stachurski, Z.; Breen, J.; Berges, H.; Spielmeyer, W. Repeat-length variation in a wheat cellulose synthase-like gene is associated with altered tiller number and stem cell wall composition. *J. Exp. Bot.* **2017**, *68*, 1519–1529. [[CrossRef](#)] [[PubMed](#)]
98. Mossé, J.; Huet, J.C.; Baudet, J. The amino acid composition of wheat grain as a function of nitrogen content. *J. Cereal Sci.* **1985**, *3*, 115–130. [[CrossRef](#)]
99. STATISTICA, Version 14; StatSoft Inc.: Tulsa, OK, USA, 2020.
100. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)]
101. Yu, J.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **2006**, *17*, 155–160. [[CrossRef](#)]
102. 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. [[CrossRef](#)]
103. Turner, S. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **2018**, *3*, 731. [[CrossRef](#)]
104. Alaux, M.; Rogers, G.; Letellier, T.; Flores, R.; Alfama, F.; Pommier, C.; Mohellibi, N.; Durand, S.; Kimmel, E.; Michotey, C.; et al. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.* **2018**, *19*, 111. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.