

Article

Nutritional Genomic Approach for Improving Grain Protein Content in Wheat

Tania Kartseva ¹, Ahmad M. Alqudah ² , Vladimir Aleksandrov ¹ , Dalia Z. Alomari ³ , Dilyana Doneva ¹, Mian Abdur Rehman Arif ⁴ , Andreas Börner ⁵  and Svetlana Misheva ^{1,*} 

¹ Institute of Plant Physiology and Genetics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 21, 1113 Sofia, Bulgaria; tania_karceva@abv.bg (T.K.); aleksandrov@gbg.bg (V.A.); donevadiliana@gmail.com (D.D.)

² 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

³ Department of Clinical Nutrition and Dietetics, Faculty of Applied Medical Sciences, The Hashemite University, P.O. Box 330127, Zarqa 13133, Jordan; daliaz.alomari@hu.edu.jo

⁴ Wheat Breeding Group, Plant Breeding and Genetics Division, Nuclear Institute for Agriculture and Biology (NIAB), Faisalabad 38000, Pakistan; m.a.rehman.arif@gmail.com

⁵ Leibniz Institute of Plant Genetics and Crop Plants Research (IPK Gatersleben), Corrensstraße 3, OT Gatersleben, 06466 Seeland, Germany; boerner@ipk-gatersleben.de

* Correspondence: slandjeva@gmail.com

Abstract: Grain protein content (GPC) is a key aspect of grain quality, a major determinant of the flour functional properties and grain nutritional value of bread wheat. Exploiting diverse germplasms to identify genes for improving crop performance and grain nutritional quality is needed to enhance food security. Here, we evaluated GPC in a panel of 255 *Triticum aestivum* L. accessions from 27 countries. GPC determined in seeds from three consecutive crop seasons varied from 8.6 to 16.4% (11.3% on average). Significant natural phenotypic variation in GPC among genotypes and seasons was detected. The population was evaluated for the presence of the trait-linked single nucleotide polymorphism (SNP) markers via a genome-wide association study (GWAS). GWAS analysis conducted with calculated best linear unbiased estimates (BLUEs) of phenotypic data and 90 K SNP array using the fixed and random model circulating probability unification (FarmCPU) model identified seven significant genomic regions harboring GPC-associated markers on chromosomes 1D, 3A, 3B, 3D, 4B and 5A, of which those on 3A and 3B shared associated SNPs with at least one crop season. The verified SNP–GPC associations provide new promising genomic signals on 3A (SNPs: *Excalibur_c13709_2568* and *wsnp_Ku_c7811_13387117*) and 3B (SNP: *BS00062734_51*) underlying protein improvement in wheat. Based on the linkage disequilibrium for significant SNPs, the most relevant candidate genes within a 4 Mbp-window included genes encoding a subtilisin-like serine protease; amino acid transporters; transcription factors; proteins with post-translational regulatory functions; metabolic proteins involved in the starch, cellulose and fatty acid biosynthesis; protective and structural proteins, and proteins associated with metal ions transport or homeostasis. The availability of molecular markers within or adjacent to the sequences of the detected candidate genes might assist a breeding strategy based on functional markers to improve genetic gains for GPC and nutritional quality in wheat.

Keywords: association mapping; bread wheat; candidate genes; micronutrients; grain proteins



Citation: 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. <https://doi.org/10.3390/foods12071399>

Academic Editor: Michael Tilley

Received: 27 February 2023

Revised: 21 March 2023

Accepted: 23 March 2023

Published: 25 March 2023



Copyright: © 2023 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

Bread wheat (*Triticum aestivum* L.) is the staple crop for an estimated 35–40% of the world population and provides more calories and protein in the human diet than any other crop [1]. Grain proteins are a major determinant of the end-use quality and an important nutritional trait of wheat [2]. Besides being a primary source of protein, wheat is also a

major supply of other essential and beneficial compounds for human health including starch, minerals, vitamins, and fibers. Hence, wheat is a decisive crop for global food security, and increases in grain production are required to feed the expanding human population. Tremendous gains in wheat production have occurred since the 1960s due to the introduction of semi-dwarf, lodging-resistant, and nitrogen-responsive wheat lines [3]. However, the growing demands of consumers for healthier and higher-nutritional-value food have shifted the focus of breeding programs on improving grain quality in addition to high productivity. Moreover, malnutrition is recognized as a serious global health issue, due to the deficiency in proteins and micronutrients. The ongoing appeal for optimal nutrition is also a driver for the food industry. For a while, innovative fortified wheat such as bakery products with improved protein and enhanced overall nutritional contents are being introduced as a short-term solution [4,5]. However, the world is looking for a sustained science-based solution relying on genetic biofortification to address the human nutritional requirements [6–8]. Such nutritional genetic approaches necessitate the identification of responsible genes and their deployment for developing improved crops [9].

Mature wheat grain contains 8 to 20% proteins (10–12% on average) [10]. Grain proteins are localized in the starchy endosperm, aleurone layer, seed coat and embryo, and are classified into three groups by their function: storage proteins (gliadins, glutenins, and certain albumins and globulins), structural and metabolic proteins, and protective proteins [11]. Grain protein content (GPC) is a quantitative trait predisposed by a complex interplay between genes and environment. A large number of genomic regions or quantitative trait loci (QTL) underlying GPC had been reported on many chromosomes. Often, narrow genetic resources and the high dependence of GPC on environmental factors are the main drawbacks that limit the detection of stable GPC-QTLs. Most of what we know about the genetic architecture of GPC relies on bi-parental populations (reviewed in [12]). In bread wheat, for example, thirteen QTL on chromosomes 2A, 2B, 2D, 3D, 4A, 6B, 7A, and 7D have been reported in a mapping population of 100 recombinant inbred lines (RILs) derived from a cross between a low-protein and a high-protein parents [13]. Other studies focused on durum wheat to explore the genetic factors of GPC. For instance, on the basis of a doubled haploid Canadian durum wheat population consisting of 162 lines derived from Pelissier × Strongfield cross and 1212 polymorphic single nucleotide polymorphisms (SNPs), four QTL (*QGpc.spa-2B.1*, *QGpc.spa-2B.2*, *QGpc.spa-3A.3*, and *QGpc.spa-7A*) have been consistently detected in multiple environments [14].

In contrast, very few of the research on GPC genetics is based on genome-wide association studies (GWAS). Association mapping is a complementary approach to bi-parental linkage analysis and searches for functional polymorphisms at many loci across a set of diverse germplasm providing higher mapping resolution [15]. GWAS has been used to identify the genetic factors behind the natural variation in different diverse populations including tetraploid and hexaploid wheat [16–19]. For instance, GWAS revealed novel genomic regions associated with high GPC in 161 wild emmer wheat lines on chromosomes 2B (*QGpc.cd1-2B.1*) and 7B (*QGpc.cd1-7B.2*) using 13,116 DArT-seq markers [16]. Using a population of around 800 spring bread wheat accessions genotyped by 5095 SNPs, significant markers associated with GPC were detected on chromosomes 2A, 3B, 3D, 5B, 6B, and 7B [17]. Nine significant GPC-associated markers had been physically mapped to chromosome 6A in 93 spring wheat cultivars using GWAS with 9351 SNPs [18]. Therefore, multi-environment screening of new large populations is an imperative need to detect novel stable QTLs. Exploring new GPC genomic regions and their deployment to enhance grain protein is one of the most effective approaches for wheat quality breeding.

Additionally, many conventional QTL studies or GWAS suggest that the accumulation of protein and other components essential for human health such as minerals are sharing common genomic regions. This may contribute to their more efficient manipulation for improving wheat grain quality, thereby mitigating global micronutrient and protein malnutrition [9,19]. For instance, the wild emmer allele *Gpc-B1* on chromosome 6B derived from *T. dicoccoides* is effective in improving grain protein, Zn and Fe concentrations [20]. QTLs con-

ferring protein content that share loci with high Zn and/or Fe have been reported on wheat chromosomes including 2A, 2B, 5A, 6B and 7B [21,22]. Such pleiotropic regions harboring co-located QTL for grain protein and micronutrient contents were recently identified on chromosomes 3A, 3B, 5A and 7D [9,23,24].

The present study aims to characterize the natural phenotypic variation for GPC in field-grown samples in a diverse panel of 255 bread wheat accessions originating from 27 countries in 5 continents, and to identify genomic regions and candidate genes associated with GPC through GWAS. Here, we found high natural variation in GPC in wheat plants and the most promising genomic intervals underlying the variation. Furthermore, we discussed the findings in the context of the genetic relations between grain protein content and micronutrient accumulation in wheat grains. Linking protein content of wheat grain with responsible genomic regions and their candidate genes will provide an extensive insight into the potential route of improving GPC and the overall nutritional quality of wheat.

2. Materials and Methods

2.1. Plant Material

A worldwide diverse collection of 255 winter wheat (*Triticum aestivum* L.) accessions was used for GPC and association mapping analyses. The population consisted of 192 advanced cultivars, 53 breeding lines, and 10 doubled haploid lines, originating from 27 countries on 5 continents (Table S1). One part of the entries was selected based on preexisting information about their performance under different growing conditions during wintertime. In addition, representatives of the core collection of the Institute of Field and Vegetable Crops (IFVCNS), Novi Sad, Serbia, and parental lines of Western European breeding programs were also included. Seed samples were provided by the Seed Genebank, Leibniz Institute for Plant Genetics and Crop Plants Research (IPK, Gatersleben), Germany. The seed material was harvested from a three-growing-season field experiment (harvest 2016, 2017, and 2018) performed in Gatersleben, Germany (51°49' N, 11°16' E), in a random design in double rows and two replications for each accession and each growing season. Information on the monthly weather statistics for the region of the experimental field is given in Figure S1.

2.2. Preparation of Grain Samples and Measurement of GPC

Seeds were dried to a constant weight and ground into fine powder with IKA Tube Mill Control (IKA Werke GmbH & Co, Staufen, Germany). For each accession and harvest year, three whole-grain powder samples (1.0 g each) were measured. The total nitrogen (N) concentration was determined via the Kjeldahl method using the UDK 159 Automatic Kjeldahl Nitrogen Protein Analyzer (Velp Scientifica, Usmate Velate, Italy). GPC (%) was calculated by multiplying grain N concentration by 5.62 according to [25].

2.3. Statistical Analysis

The phenotypic data were analyzed (by ANOVA) to check the significant differences in GPC among accessions, environments (growing seasons) and to test the interaction effects. The best linear unbiased estimates (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 using the fixed and random model circulating probability unification (FarmCPU).

Pearson's correlation coefficients (r) were computed to test the relationships for GPC among the growing seasons and with the BLUEs.

To measure the genetic influences of GPC, the broad-sense heritability (H^2) was estimated according to the following formula:

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

where σ_G^2 is the genotype variance, σ_E^2 is the variance of the residual and nE is the number of the of environments (growing seasons).

To estimate the heritability of GPC in each environment, the same formula was used, where σ_G^2 is the genotype variance, σ_E^2 is the variance of the residual, but nE is the number of the replications in a given environment.

All phenotypic data analyses were accomplished using STATISTICA 14 [26].

2.4. Phylogenetic Analysis

The wheat panel has been previously genotyped using the 90 K iSELECT SNP array by the SGS Institut Fresenius GmbH TraitGenetics Section, Gatersleben, Germany. To perform phylogenetic analysis, a dataset of 17,093 polymorphic SNPs for the 255 wheat accessions was used as input data in the R package “phyclust” [27]. The phylogenetic tree was constructed by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm [28].

2.5. Genome-Wide Association Study (GWAS) Analysis

The population was previously described and more information about the filters applied to the SNP markers database, population structure, and linkage disequilibrium (LD) is provided in [29].

GWAS was performed by applying the FarmCPU algorithm through GAPIT 3 (Genomic Association and Prediction Integrated Tool) in R. This algorithm was implemented to avoid any false-negative associations and to control the false positive ones in accordance with [15]. The association signals by GWAS were calculated for phenotypic data of each growing season and BLUEs. On the resulting Manhattan plots, the threshold of statistically significant QTNs (quantitative trait nucleotides) or MTAs (marker-trait associations) was set above the false discovery rate (FDR) at the significance level of 0.01 ($-\log_{10} p\text{-values} > \text{FDR}$) calculated for each growing season separately as per [15]. Important p -value distributions (expected vs. observed $-\log_{10} (p\text{-values})$) were shown with a quantile–quantile plot.

2.6. Candidate Gene Search

LD was calculated and defined by Schierenbeck et al. [29], and was used here as interval for detecting the candidate gene(s). High-confidence putative candidate genes were predicted based on significant SNPs, which passed the FDR threshold, and their extension regions within the LD interval (± 2 Mbp) using the reference genome sequence of cv. Chinese Spring by blasting against IWGSC RefSeq annotation v1.1. [30]. Three web-based platforms (WheatIs, EnsemblPlants, and Persephone) were used to obtain the gene annotations, gene ontologies (GO) and InterPros, and descriptions for the potential candidate genes (<https://urgi.versailles.inrae.fr/wheatIs/>; http://plants.ensembl.org/Triticum_aestivum/Info/Index; <https://web.persephonesoft.com/?data>) (accessed on 10 June 2022).

3. Results

3.1. Phenotypic Assessment

Large phenotypic variation was observed for GPC across the whole dataset from three crop growing seasons within the panel of 255 wheat accessions (Figure 1A,B, Table S1). The GPC values presented as a percentage (%) in the accessions for the individual growing seasons ranged from 5.7 to 18.1%. The accession values averaged over the three grow-

ing seasons ranged from 8.6 to 16.4% with a total average of $11.3\% \pm 0.09$. The widest phenotypic variation of GPC was observed in 2016 growing season followed by 2017 (Figure 1B). To assess the population phenotypic variability across the growing seasons, we calculated the deviations from the corresponding annual average GPC (in %); then, for each accession, the deviations over the three seasons were averaged. Genotypic effects varied along with the environmental gradient; the variance ranged from 0.024 (2018) to 0.031 (2017) (Table S2). The population variance over the growing seasons ranged from 0 to 0.147 (Table S2). Some accessions displayed consistently high GPC across the environments, for instance, Bezenchukskaya 380, Rannaya 12, Moskovskaya 40 (Russia) and ZG K 238/82 (Croatia) (high positive deviation from the average GPC, low variance), whereas others such as Wagnein (Austria), was inconsistent across the crop seasons although with a high average GPC (high positive deviation from the average GPC, high variance), or had a consistently low GPC, for instance, Folke (Sweden), Durin and GSA 3 (France) (high negative deviation from the average GPC, low variance) (Table S2).

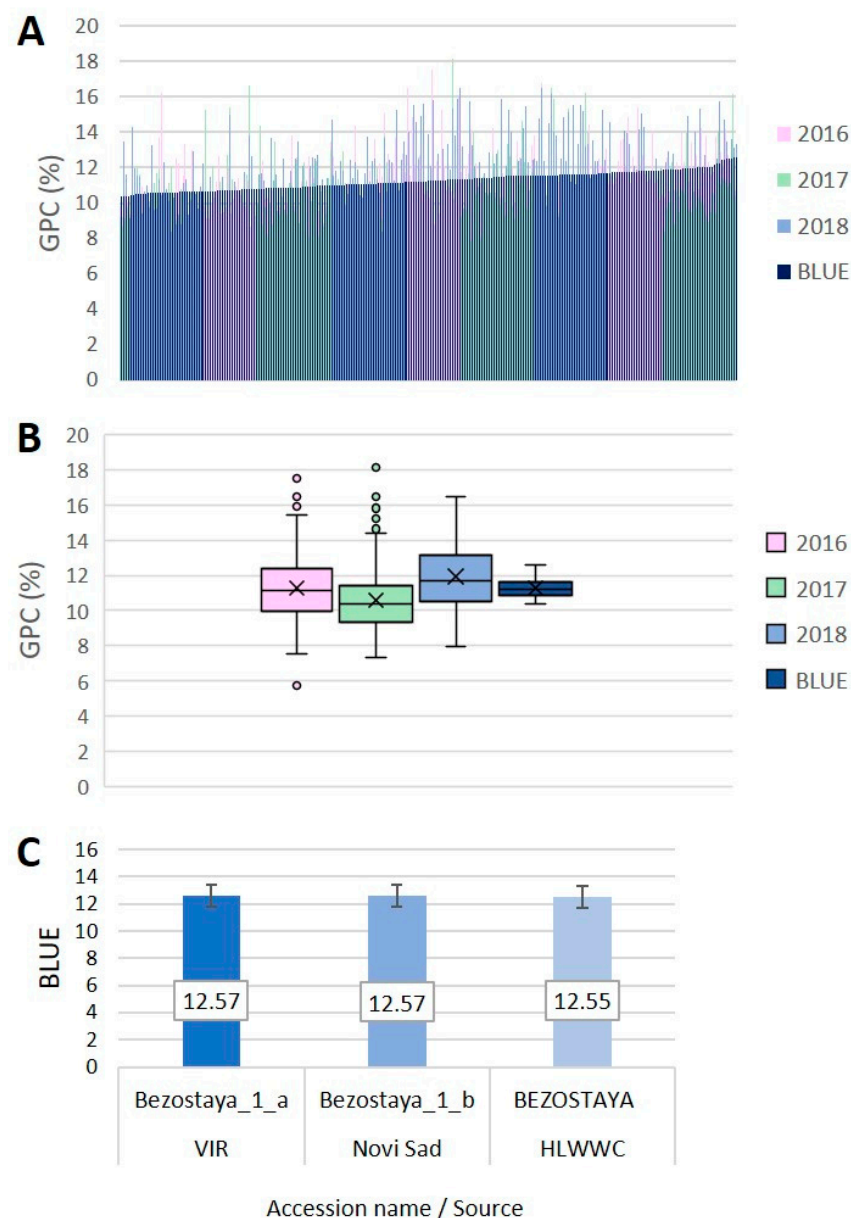


Figure 1. (A) Distribution of 255 bread wheat accessions from 27 countries for grain protein content (GPC, %); (B) box-plots of GPC (%) in three growing seasons (2016, 2017, 2018), and best linear unbiased estimates (BLUEs); (C) top three best accessions based on BLUEs.

The factorial analysis of variance showed significant effects of Genotype, Growing season and their interaction (Table S3). To eliminate the environmental impact, BLUEs were calculated for each accession across the growing seasons by assuming the genotype as a fixed effect and the growing season as a random effect. The distribution of the BLUEs across growing seasons also showed wide genotypic variation in the range 10.4–12.6% (Figure 1A,B, Table S1). The three accessions with the highest average GPC based on BLUEs were all samples of the Russian cultivar Bezostaya 1 coming from different sources (genebank collections) (Figure 1C). Accessions from Northern, Western and most of Central Europe, as well as Canada had average protein values lower than the BLUE averaged over the entire panel, whereas accessions from Eastern Europe, part of Central Europe (Hungary and Austria), Asia, and Australia had average protein values higher than the average BLUE (Table S4).

As expected for a quantitative trait, the distribution of GPC means of the 255 accessions followed normal curves across the three growing seasons, with a large number of accessions covering protein contents within the range of 10–12% for each environment (Figure 2A). Low to moderate positive Pearson's correlation coefficients, denoted as r , were calculated for GPC across the growing seasons and with the BLUEs (Figure 2B).

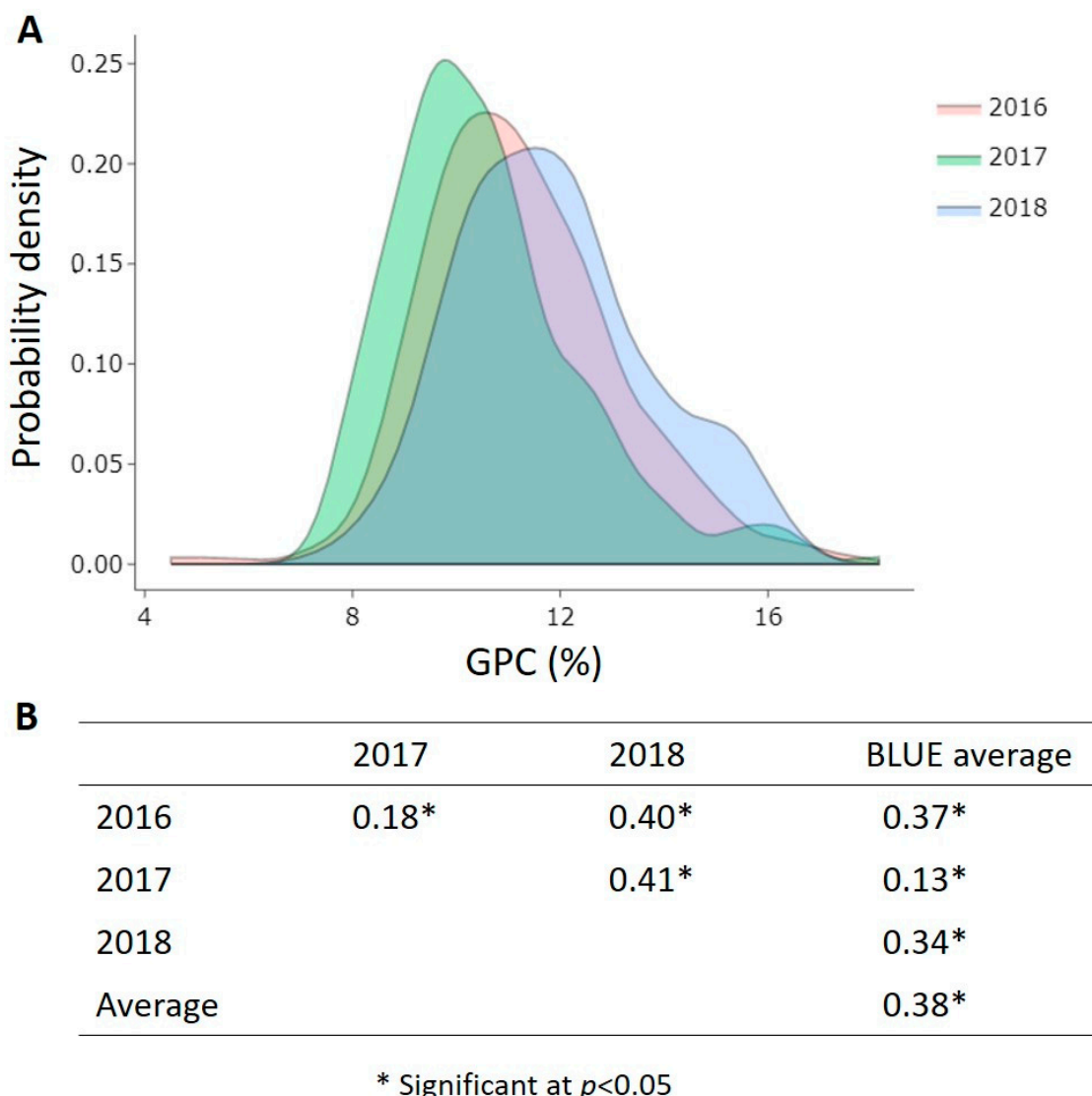


Figure 2. (A) Probability density of 255 bread wheat accessions from 27 countries for GPC (%) across three growing seasons (2016, 2017, 2018); (B) Pearson correlation coefficient values r of GPC among all measured environments and the BLUEs.

All genotypes were grouped into three clusters of low (mean 10.7%), medium (mean 11.2%) and high (mean 11.8%) GPC based on BLUEs, and comprising 77, 80, and 98 accessions, respectively (Figure 3A, Table S1). Figure 3B shows the proportion of the entries of low-, medium-, and high-grain protein in the pools from each country.

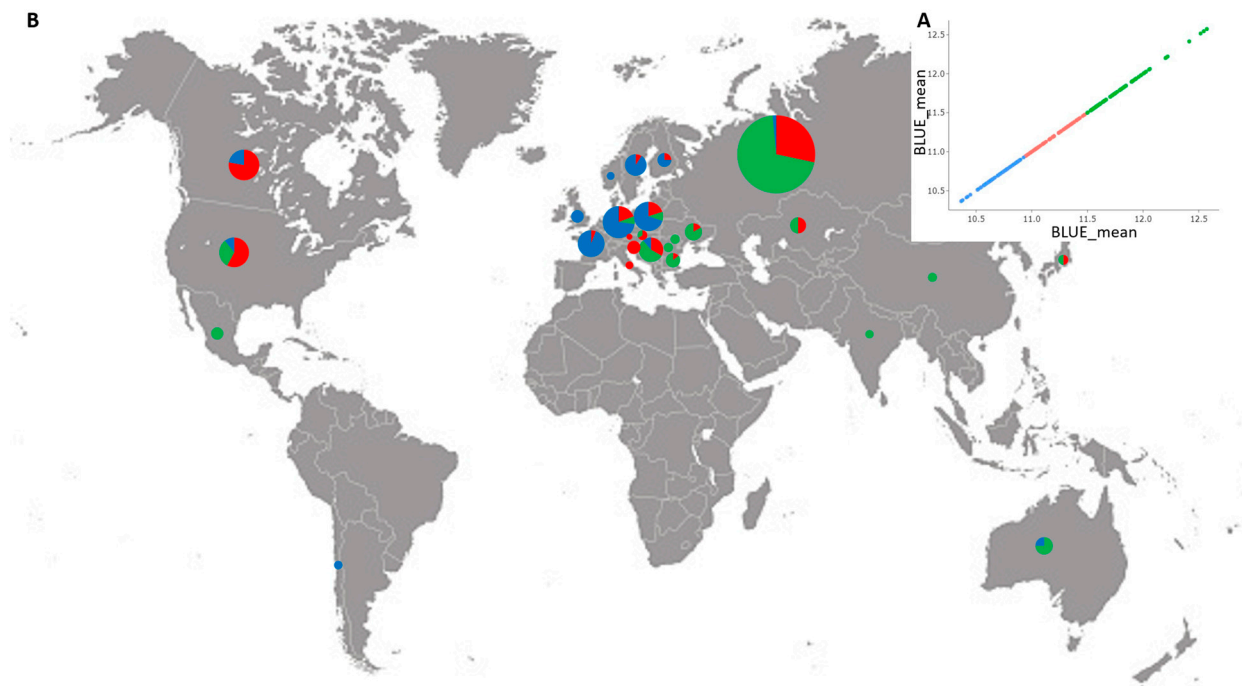


Figure 3. (A) Cluster distribution of 255 bread wheat accessions from 27 countries for GPC based on BLUEs averaged over three growing seasons; (B) geographic distribution of the wheat accessions. Circles indicate the proportion of accessions with low (blue), medium (red), and high (green) GPC in each country pool.

3.2. Phylogenetic Analysis

The phylogenetic tree analysis of the 255 diverse wheat accessions was based on 17,093 polymorphic SNP markers. The accessions were grouped into six distinctive clusters related mainly to their geographical origin (Figure S2). The largest cluster (I) consisted of 99 accessions predominantly from the former Soviet Union (Russia, Kazakhstan and Ukraine), few accessions from Bulgaria, and a sub-cluster of US cultivars. Clusters III and V comprised almost exclusively accessions from North America and Scandinavia, respectively. The accessions from South-Eastern Europe formed the core of cluster IV, whereas those from Central and Western Europe dominated in cluster VI. Cluster II comprised 16 accessions of mixed origin.

3.3. Association Mapping Analysis

For GWAS, we used the genotype data for the 255 accessions that were already available from the 90 K iSELECT SNP technology (SGS Institut Fresenius GmbH TraitGenetics Section, Gatersleben, Germany), and described by Schierenbeck et al. [29]. Although high broad-sense heritability estimates were observed for GPC in the individual growing seasons ($H^2 = 0.79$ in 2016, 0.68 in 2017, and 0.78 in 2018) (Table S1), on average $H^2 = 0.79$, the analysis of variance revealed significant genotype \times environment interaction (Table S3). Therefore, we performed GWAS for GPC first for each growing season, followed by analysis based on BLUEs.

In total, 22 significant QTNs or MTAs for GPC were detected across the three growing seasons and based on BLUEs at $-\log_{10}(p\text{-value}) > \text{FDR}$ (Table 1, Figure 4A). The FDR threshold was at $-\log_{10} > 4.2$, and therefore, the number of spurious associations was

reduced. The 22 significant QTNs were located on 14 chromosomes: 1B (1 in 2018), 1D (1 with BLUEs), 2B (1 in 2016), 3A (2 with BLUEs, 2 in 2016, 2 in 2017, 1 in 2018), 3B (1 with BLUEs, 1 in 2017, 1 in 2018), 3D (1 with BLUEs), 4A (1 in 2018), 4B (1 with BLUEs), 5A (1 with BLUEs), 5B (1 in 2018), 6A (1 in 2016), 6B (1 in 2018), 7A (1 in 2018) and 7B (1 in 2016) (Table 1, Figure 4A). Most of the QTNs showed negative additive effect reducing the GPC value (ranging from -0.86 to -0.30%). The findings highlighted four SNPs (*Excalibur_rep_c112060_100* on 7B, *Excalibur_c39248_485* and *Excalibur_c13709_2568* on 3A, and *wsnp_Ex_c16432_24932860* on 5B) with positive additive effect ranging from 0.32 to 0.70% which can be used to increase GPC.

Table 1. Significant single nucleotide polymorphism (SNP) markers for grain protein content (GPC) in a diverse bread wheat population of 255 accessions, identified by the fixed and random model circulating probability unification (FarmCPU) model at the threshold of $-\log_{10}(p\text{-value}) > \text{FDR}$ (False Discovery Rate).

	SNP	Chr.	Position (bp)	Allele	Effect	p-Value	$-\log_{10}$
GPC_BLUE							
1.	<i>Ex_c67198_1838</i>	1D	10,259,568	T/C	-0.385547705	2.42×10^{-6}	5.616184634
2.	<i>Excalibur_c13709_2568</i>	3A	518,897,423	A/G	0.543011247	4.20×10^{-9}	8.37675071
3.	<i>wsnp_Ku_c7811_13387117</i>	3A	714,304,435	A/G	-0.391805162	2.43×10^{-7}	6.614393726
4.	<i>BS00062734_51</i>	3B	545,069,014	A/G	-0.323156108	1.89×10^{-6}	5.723538196
5.	<i>Kukri_c11944_436</i>	3D	609,621,771	T/C	-0.318723447	6.13×10^{-6}	5.212539525
6.	<i>wsnp_Ra_c1146_2307483</i>	4B	630,470,822	T/G	-0.304930639	3.38×10^{-6}	5.4710833
7.	<i>BobWhite_c6759_365</i>	5A	488,262,559	T/C	-0.354037284	3.39×10^{-7}	6.469800302
GPC_2018							
8.	<i>Excalibur_c13709_2568</i>	3A	518,897,423	A/G	0.570293042	8.48×10^{-6}	5.071604148
9.	<i>BS00062734_51</i>	3B	545,069,014	A/G	-0.323156108	2.33×10^{-6}	5.632644079
10.	<i>Kukri_c37738_417</i>	1B	5,519,895	T/C	-0.570	8.48×10^{-6}	5.632644
11.	<i>Tdurum_contig46583_2146</i>	4A	738,781,353	T/C	-0.418	2.33×10^{-6}	5.071604
12.	<i>wsnp_Ex_c16432_24932860</i>	5B	483,000,817	T/C	0.696768643	6.76×10^{-7}	6.170053304
13.	<i>Excalibur_c28759_914</i>	6B	716,259,097	A/G	-0.445651367	1.83×10^{-5}	4.73754891
14.	<i>wsnp_Ex_c22547_31738007</i>	7A	694,546,597	T/C	-0.324192625	3.17×10^{-5}	4.498940738
GPC_2017							
15.	<i>wsnp_BE490651A-Ta_2_2</i>	3A	565,466,354	T/C	-0.861806961	7.35×10^{-6}	5.133712661
16.	<i>wsnp_Ku_c7811_13387117</i>	3A	714,304,435	A/G	-0.478132901	3.64×10^{-5}	4.438898616
17.	<i>BS00062734_51</i>	3B	545,069,014	A/G	-0.421626697	4.32×10^{-5}	4.364516253
GPC_2016							
18.	<i>Excalibur_c12675_1789</i>	2B	788,655,682	A/G	-0.402444191	4.57×10^{-7}	6.3400838
19.	<i>Ku_c56370_1155</i>	3A	521,719,100	A/G	-0.355731398	7.96×10^{-6}	5.099086932
20.	<i>Excalibur_c39248_485</i>	3A	624,946,759	A/G	0.377677627	3.07×10^{-5}	4.512861625
21.	<i>Kukri_c8274_502</i>	6A	297,714,501	A/G	-0.726532079	9.30×10^{-7}	6.031517051
22.	<i>Excalibur_rep_c112060_100</i>	7B	646,065,982	T/C	0.316001498	2.26×10^{-5}	4.645891561

Three shared marker associations were found: two on chromosome 3A (markers *Excalibur_c13709_2568*, detected in growing season 2018 and based on BLUEs, and *wsnp_Ku_c7811_13387117*, detected in 2017 and with BLUEs), and one on chromosome 3B (*BS00062734_51*, detected in growing seasons 2017 and 2018, and based on BLUEs) (Table 1, Figure 4A). These QTNs could be considered as stable ones. Although no shared SNP was found with growing season 2016, a significant marker (*Ku_c56370_1155*) was detected on chromosome 3A at a position very close (2.8 Mbp) to the stable marker *Excalibur_c13709_2568*. The common SNPs showed contrasting effects on the GPC. The marker *Excalibur_c13709_2568* had positive influence that can be used to improve the protein content in wheat grains. In contrast, the two other shared SNPs showed a negative effect on GPC (Table 1).

Applying FarmCPU model using 90 K SNPs for the association analyses, we detected strong association signal within several interesting genomic regions. The Q-Q plots supported these findings and therefore, we used significant SNPs/QTNs that were obtained

with BLUEs and verified in at least one crop season in further analyses for detecting candidate genes.

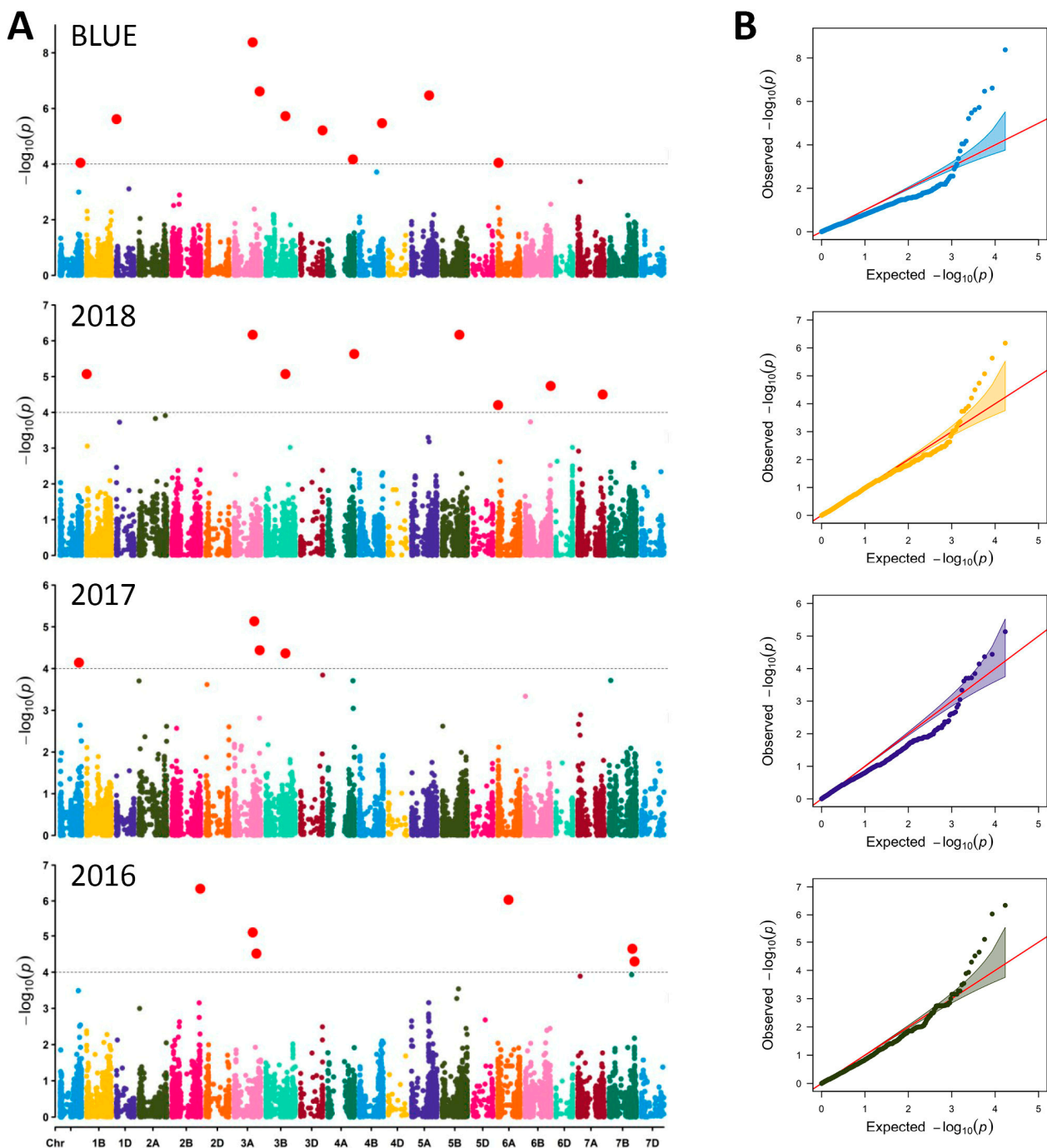


Figure 4. (A) Manhattan plots of genome-wide association study for GPC in 255 bread wheat accessions in three growing seasons (2016, 2017, 2018) and BLUEs based on FarmCPU model. The grey color line in the figure corresponds to the threshold of $-\log_{10}(p < 0.001) = 4.2$. The significant quantitative trait nucleotides (QTNs) are above the grey color line; (B) quantile–quantile plots of GPC; red line depicts the expected values.

3.4. Candidate Genes Associated with Significant SNPs

Applying the criteria to select the most significant SNPs/QTNs, 127 high-confidence genes in total were found based on the markers' physical position and within the LD interval (± 2 Mbp) determined by the LD decay. There were 19, 74 and 34 candidate

genes, located within the LD screen of the three significant shared QTNs on chromosomes 3A (markers *Excalibur_c13709_2568* and *wsnp_Ku_c7811_13387117*), and 3B (marker *BS00062734_51*), respectively. The found putative candidate genes included 122 genes, 9 probable transposons, 1 pseudogene; 4 were undefined. Of these, 113 were indicated to have molecular functions, 61 were biological processes, and 13 were involved in cellular components, as shown in Figure 5. Detailed descriptions of these candidate genes are summarized in Table S5.

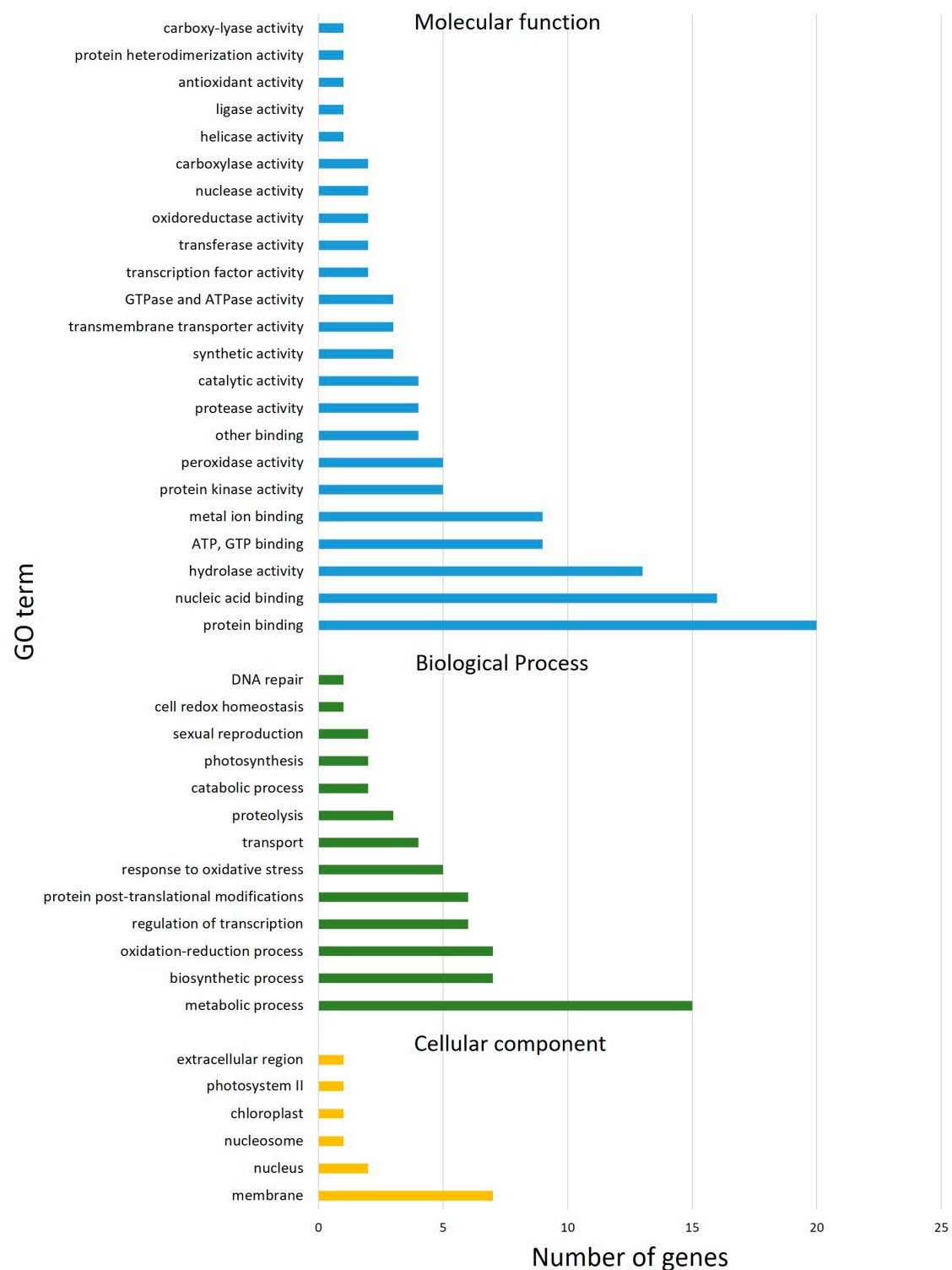


Figure 5. Gene ontology (GO) enrichment analysis.

The top five ontologies attributed to molecular function were: protein binding, nucleic acid binding, hydrolase activity, ATP and GTP binding, and metal ion binding. Similarly, for biological processes, the top five categories were: metabolic process (including carbohydrate, DNA, and amino-acids metabolism), biosynthetic process, oxidation-reduction process, regulation of transcription, and protein post-translational modifications. The key cellular component where these molecular function and biological process ontologies functionally comprised the biological system were membranes and their integral constituents. The complete list of molecular function, biological process, and cellular component ontologies is tabulated in the Supplementary Information (Table S5).

Most of the genes related to molecular functions and biological processes which were identified through GO analysis, code for proteins that were categorized into the following classes: (1) transporter proteins; (2) transcription factors; (3) proteins involved in post-translational modifications; (4) proteins involved in biosynthesis of macromolecules; (5) protective proteins; (6) structural proteins; (7) proteins bound up in metal ions transport and homeostasis. Putative key genes pertaining to different processes related to the accumulation of protein during grain development that possibly are differentially expressed in the population are presented in Table 2, and the most prominent ones are discussed in the next section.

Table 2. Potential candidate genes underlying GPC based on IWGSC RefSeq annotation v1.1.

Gene ID	Chr./Position	GO Ontology	Description
Transporter proteins			
<i>TraesCS3A02G484600</i> <i>TraesCS3A02G484900</i>	chr3A:714289919..714291827 (–strand) chr3A:714415263..714423339 (–strand)	BP: amino acid transmembrane transport; MF: amino acid transmembrane transporter activity; CC: membrane	Amino acid permease
Transcription factors			
<i>TraesCS3A02G485400</i> <i>TraesCS3A02G485500</i> <i>TraesCS3A02G486500</i>	chr3A:7145774421..714575250 (+strand) chr3A:714577793..714578643 (+strand) chr3A:714650271..714650843 (–strand)	MF: DNA binding; BP: regulation of transcription, DNA-templated	NAC domain-containing protein
Proteins, involved in post-translational modifications			
<i>TraesCS3A02G290300</i> <i>TraesCS3A02G485600</i> <i>TraesCS3A02G486400</i>	chr3A:519244293..519255598 (–strand) chr3A:714588344..714590684 (–strand) chr3A:714643608..714646388 (+strand)	MF: protein kinase activity; MF: ATP binding; BP: protein phosphorylation	Receptor-like kinase Serine/threonine-protein kinase
<i>TraesCS3A02G488100</i>	chr3A:716067049..716070839 (+strand)	BP: protein glycosylation; MF: acetylglucosaminyltransferase activity	Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
<i>TraesCS3A02G481900</i> <i>TraesCS3A02G487000</i> <i>TraesCS3A02G484000</i> <i>TraesCS3A02G483900</i> <i>TraesCS3A02G482600</i> <i>TraesCS3A02G485800</i> <i>TraesCS3A02G487100</i> <i>TraesCS3B02G340400</i>	chr3A:712621085..712621999 (+strand) chr3A:714787620..714789640 (+strand) chr3A:713962301..713963870 (–strand) chr3A:713945850..713955055 (–strand) chr3A:713210615..713212190 (+strand) chr3A:714595534..714599278 (+strand) chr3A:714939243..714940190 (+strand) chr3B:547054749..547060421 (–strand)	MF: protein binding	E3 ubiquitin-protein ligase E2 ubiquitin-conjugating enzyme
* <i>TraesCS3B02G339000</i>	chr3B:545066258..545069534 (+strand)	MF: catalytic activity; BP: lipoate biosynthetic process; CC: chloroplast; MF: lipoate synthase activity; MF: iron-sulfur cluster binding; MF: 4 iron, 4 sulfur cluster binding	Lipoyl synthase
Proteins involved in biosynthesis of macromolecules			
<i>TraesCS3A02G289800</i>	chr3A:518720116..518723505 (+strand)		PROTEIN TARGETING TO STARCH (PTST)
* <i>TraesCS3A02G484800</i>	chr3A:714298564..714310707 (–strand)	MF: catalytic activity; MF: acetyl-CoA carboxylase activity; MF: biotin carboxylase activity; MF: ATP binding; BP: fatty acid biosynthetic process; MF: ligase activity; MF: metal ion binding	Acetyl-CoA carboxylase
* <i>TraesCS3A02G289900</i> <i>TraesCS3B02G337100</i>	chr3A:518895097..518901097 (+strand) chr3B:543466729..543471191 (–strand)	CC: membrane; MF: cellulose synthase activity; BP: cellulose biosynthetic process	Cellulose synthase

Table 2. Cont.

Gene ID	Chr/Position	GO Ontology	Description
Protective proteins			
TraesCS3A02G483800	chr3A:713930115..713932275 (−strand)	MF: hydrolase activity, hydrolyzing O-glycosyl compounds; BP: carbohydrate metabolic process	Beta-1,3-glucanase
TraesCS3A02G483700	chr3A:713867934..713870026 (−strand)		
TraesCS3A02G483600	chr3A:713849212..713849871 (−strand)		
TraesCS3A02G483100	chr3A:713622221..713623907 (−strand)		
TraesCS3A02G483000	chr3A:713534311..713536049 (−strand)		
TraesCS3A02G482200	chr3A:712728151..712729578 (−strand)		
TraesCS3A02G482000	chr3A:712698902..712700166 (+strand)		
TraesCS3A02G481600	chr3A:712552546..712554003 (+strand)		
TraesCS3A02G481500	chr3A:712545116..712546522 (+strand)		
TraesCS3A02G485200	chr3A:714535163..714536122 (+strand)		
TraesCS3A02G486000	chr3A:714614717..714616126 (+strand)	MF: protein binding	Glutathione S-transferase
TraesCS3A02G487700	chr3A:715593874..715594918 (−strand)		
TraesCS3A02G488200	chr3A:716097184..716098213 (−strand)	MF: protein kinase activity; MF: ATP binding; BP: protein phosphorylation; MF: carbohydrate binding	Lectin receptor kinase, Protein kinase domain; Serine/threonine-protein kinase; Concanavalin A-like lectin/glucanase domain
TraesCS3A02G488300	chr3A:716105661..716106534 (−strand)		
TraesCS3B02G338900	chr3B:545061480..545063744 (+strand)	MF: peroxidase activity; BP: response to oxidative stress; MF: heme binding; BP: oxidation-reduction process	Peroxidase
TraesCS3B02G338800	chr3B:545048539..545050613 (+strand)		
TraesCS3B02G339400	chr3B:545385674..545385674 (+strand)		
TraesCS3B02G339500	chr3B:545432618..545433981 (+strand)		
TraesCS3B02G339600	chr3B:545493643..545495202 (+strand)		
TraesCS3B02G339700	chr3B:545562324..545563678 (+strand)		
TraesCS3B02G339800	chr3B:545667818..545669160 (+strand)	MF: serine-type endopeptidase activity; BP: proteolysis	Subtilisin-like protease
TraesCS3B02G339100	chr3B:545179979..545187120 (+strand)		
Structural proteins			
TraesCS3A02G484500	chr3A:714223433..714224075 (+strand)	CC: nucleosome; MF: DNA binding; MF: protein heterodimerization activity	Histone H3
TraesCS3B02G339900	chr3B:5457066521..545709391 (+strand)	CC: nucleus; BP: regulation of transcription, DNA-templated	40S ribosomal protein S25
TraesCS3A02G488000	chr3A:715973684..715975189 (+strand)	CC: extracellular region; BP: sexual reproduction	Expansin protein
Proteins involved in metal ions transport and homeostasis			
TraesCS3A02G484300	chr3A:714150436..714156128 (+strand)	MF: calcium ion binding	Voltage-dependent L-type calcium channel subunit
TraesCS3A02G488400	chr3A:716258725..716259719 (+strand)		Calcium binding protein
TraesCS3B02G338600	chr3B:545008333..545008942 (−strand)		
TraesCS3B02G338400	chr3B:544921860..544922297 (−strand)		
TraesCS3B02G338200	chr3B:544818884..544819617 (−strand)		
TraesCS3A02G484700	chr3A:714292993..714297477 (−strand)	BP: amino acid transmembrane transport; MF: amino acid transmembrane transporter activity; CC: membrane	Metallothionein

* Genes hit by a significant stable QTN.

4. Discussion

The grain protein content is one of the main determinants of the nutritional and baking quality of wheat. More than 80% of their final N content wheat plants accumulate before anthesis as leaf plastidial proteins [31]. During the grain-filling period, which coincides with leaf senescence, the degradation of plastidial proteins, mainly Rubisco, provides the N available for remobilization from leaves to ears to be used for protein synthesis. Due to the reverse relationship between grain yield and quality, the selection of wheat cultivars of high yield influences the protein levels in the grain despite the better economic return in areas where cereals contribute to the largest daily dietary intake. The resulting deficiency in proteins and essential micronutrients in the cereal-based diet known as “hidden hunger” is recognized as one of the severe health issues for pregnant women and young children, particularly in developing countries. Enhancing the nutritive levels of wheat plants using molecular breeding approaches relies on the knowledge of phenotypic variability and the genetic dissection of the nutrition associated traits.

4.1. Phenotypic Variation for Grain Protein Content

Growing wheat with high grain protein begins with selecting of appropriate genotypes followed by management practices [32]. Therefore, the information about genetic diversity for this trait is important. Since GPC is highly influenced by the interaction between the environmental and genetic factors, our study was designed to consider the impact of these main factors to estimate GPC using world-wide diverse genotypes across three consecutive growing seasons. In the current study, the phenotypic data analyses revealed considerable variation for GPC. The obtained high heritability and the moderate correlations observed among growing seasons demonstrated consistency of GPC across the environments and suggested a substantial genetic component to the variation of GPC. Comparable heritability estimates were also reported in previous studies [14,33].

Environmental factors, in particular reduced precipitation and rising temperatures at critical vegetation phases are a major pressure on wheat production with impact on both grain yield and GPC [34]. During the three growing seasons, the average monthly temperatures were, in general, close to or higher than the climate norm for the seed-providing experimental field (Figure S1). However, in 2018, from heading to the end of grain filling phase (May–July), the precipitation was significantly lower compared to the climate norm, coupled with higher average monthly temperatures (up to 4 °C deviation), which is suggestive of prolonged drought. During the grain filling phase (July), the precipitation was considerably reduced in 2016, whereas in 2017 extreme rainfall values were recorded (Figure S1). These fluctuations among the three crop seasons might also account for the observed variation in the protein content in the studied set of 255 wheat accessions. Terminal drought and especially the prolonged drought period have been shown to increase GPC due to elevated N uptake and proline synthesis when plants suffer from water insufficiency [34]. In our study, the average protein values for 2016 (11.29%) and 2018 (11.93%), when drought periods were suggested during the late vegetation, were significantly higher compared to 2017 (10.60%) ($p < 0.001$; Table S3). The Genotype \times Growing season interaction was significant (Table S3), and the genotypic effects varied along with the environmental gradient, in particular, variation of different extent was observed in protein levels in accessions of various geographic origin across the environments (Table S2). The results on the phenotypic variability for protein content is largely in accordance with the spatial distribution of the wheat accessions demonstrated also by the phylogenetic tree (Figure S2). The 255 diverse accessions were clustered into six distinctive groups mainly based on their geographic origin. For example, the largest cluster (I) comprised accessions from Russia, Ukraine, Kazakhstan, and Bulgaria (Figure S2). The accessions pools from these countries have large fractions of high protein entries (Figure 3). The obtained outcome is in line with the grain quality reports for the wheat germplasm from these countries [18,35–41]. In addition, we assume that the accessions grouped, to some extent, according to their genealogy. For example, based on known pedigree information, Russian cultivars especially Bezostaya 1, Avrora, Skorospelka 35, Kavkaz, Mironovskaya 808, Rannaya 12, Krasnodarskij karlik, as well as Ukrainian cultivars, such as Odesskaja 86, Obrij, Khersonskaja 552, Yuzhnaya Zarya, etc., have been broadly used as parents for hybridization by the breeders in Bulgaria since the 1950s [42]. The Russian cultivar Bezostaya 1 demonstrated the highest grain protein based on BLUEs, and this result was consistent in the three entries maintained at different seed gene banks. This partly explains the obtained high values of the entries from the aforementioned European germplasm and their positioning in the same cluster.

4.2. Quantitative Trait Nucleotides

Applying GWAS approach in the diverse population with dense SNPs using GPC values (%), empowered us to detect 15 significant QTNs distributed on 10 chromosomes, whereas GWAS performed with BLUEs identified 7 significant QTNs on 6 chromosomes. The most significant QTNs were mapped on chromosomes 1B, 1D, 2B, 3A, 3B, 3D, 4A, 4B, 5A, 5B, 6A, 6B, 7A, and 7B, which were mostly detected in single environments (crop

growing seasons). These results are in agreement with previous studies reporting marker associations with GPC using different wheat populations. Thus, using a global core collection, made up of 372 accessions from 70 countries, 33 markers spread over 15 wheat chromosomes (1A, 1B, 1D, 2A, 2B, 2D, 3B, 3D, 4B, 5B, 5D, 6A, 6B, 6D, and 7B) were detected in association with GPC [43]. The GWAS approach applied to CIMMYT breeding material identified new GPC marker associations on chromosomes 6A and 7A [44]. More recent association mapping studies revealed novel QTL regions on chromosomes 2B, 5A, 5D, and 7B in a collection of local and modern Iranian accessions [45]. In wild emmer, Liu et al. [16] identified molecular markers grouped into two novel GPC-linked QTL regions on chromosomes 2B and 7B, using a worldwide collection of 486 wheat accessions from China, Europe, and America. A recent study detected genomic regions associated with GPC on chromosomes 2B and 6A, using a diverse panel of 372 European accessions [46]. In total, 10 GPC-linked QTLs represented by 45 SNPs were mapped on chromosomes 1A, 1B, 2B, 3A, 3D and 6A, using a set of 486 accessions from worldwide regions, especially China [47]. Alemu et al. [17] reported many chromosomal regions on 2A, 3B, 3D, 5B, 6A, 6B and 7B as the source of highly significant MTAs for GPC based on GWAS analysis in 802 elite spring breeding lines. The contribution of loci on chromosomes 1A, 2A, 2B, 3A, 3B, and 6B for protein content was confirmed in a GWAS study on UK wheat genotypes [48], whereas genomic regions on chromosome 6A were recently confirmed in a study on Russian spring cultivars [18]. QTNs for GPC under limited water supply were located on chromosomes 1B, 2A, 2B, and 6B [49], of which the QTL 6B.7 was co-mapped with the previously cloned gene *Gpc-B1* associated with high grain protein [20]. In total, 16 significant MTAs for GPC on chromosomes 1A, 1B, 1D, 2B, 3B, 4B, 5A, 5B, 5D, 6A, 6B and 7A have been identified in a set of 280 diverse bread wheat genotypes [50]. QTL hotspots harboring a number of significant MTAs for protein content have been recently reported on chromosomes 3B (13), 5A (2), 5B (6) and 7D (2) in a population of Mediterranean landraces [51].

A recent meta-analysis of GPC-associated QTLs in hexaploid and tetraploid wheat conducted on 48 linkage-based mapping studies identified stable QTLs (meta-QTLs) and hotspots on all wheat chromosomes, except for 1D and 4D, of which 19 meta-QTLs and 2 hotspots on 14 chromosomes were shown to be co-localized with significant SNP-based MTAs retrieved from association mapping studies [52].

The significant QTNs identified in the current study are potentially promising for early selection to improve the marker-assisted selection of wheat grain quality. However, the potential use of marker associations may be questioned considering the variation in their stability and therefore, requires multi-environmentally verified ones [53]. Our results highlighted three stable QTNs: one on 3B that was consistent in three environments (2017, 2018 and BLUEs), one on 3A detected in 2017 and with BLUEs, and another one on 3A verified in two environments (2018 and BLUEs). No shared marker was found with crop season 2016, indicating that due to the environment, there are different genetic SNPs controlling GPC. The consistent SNP on chromosome 3B (*BS00062734_51*) is located at the position of 545,069,014 bp on the IWGSC RefSeq v1.1 map and is, therefore, distant from the two significant SNPs identified on the same chromosome in the study by Alemu et al. [17] within the intervals 822,819,158–822,820,535 bp, and 795,567,173–795,572,772 bp, as well as far from the five significant SNPs on 3B reported by Leonova et al. [18], altogether located between 29,356,100 and 493,122,600 bp. It is also far from the reported GPC-SNPs on chromosome 3B (*wsnp_Ex_c20652_29734133* at 292,024,034 bp, and *RAC875_c58159_989* at 564,248,743 bp [48], and *AX-94746929* at 800,933,346 bp [50]). The positions of the two stable SNPs, identified in the present study on chromosome 3A (*Excalibur_c13709_2568* at 518,897,423 bp and *wsnp_Ku_c7811_13387117* at 714,304,435 bp), are far from the significant GPC-QTL peaked by *wsnp_Ku_c30545_40369365* at 363,458,708 bp [48], as well as far from the two significant SNPs on 3A detected by Leonova et al. [18] in single crop seasons (located at 20,019,900 and 739,397,100 bp). In addition, both are distantly located from the multi-environmentally verified SNPs mapped to the 3A chromosome region between 191,530,793 and 484,643,825 bp [47]. Moreover, the location comparisons of the stable

QTNs identified in our study with those of the three meta-QTLs on chromosome 3A and the two GWAS-validated meta-QTLs on chromosome 3B [52] also imply that these are different loci. This suggests that chromosomes 3A and 3B harbor one or more genomic regions associated with GPC that had not been reported yet. Therefore, we can claim with reasonable certainty that the significant QTNs/SNP associations with GPC on chromosomes 3A and 3B identified in the present study are novel.

Interestingly, the most promising and stable QTNs in our study located on 3A and 3B were found to be strongly associated with other human health components such as macro- and microminerals and other grain yield components [9,19,23], indicating that these QTNs can be useful for the simultaneous improvement not only of protein contents but also for minerals and grain yield.

4.3. Candidate Genes

We selected the three significant QTNs obtained with BLUEs on chromosomes 3A and 3B, that were shared with at least one individual growing season and, thereby were considered as stable ones, to pull out putative genes that are potentially responsible for wheat GPC. Using the latest IWGSC RefSeq annotation v1.1, we retrieved 127 high-confidence GPC-related candidate genes in the vicinity (± 2 Mbp) of the stable QTNs. Below, we discuss some of the most prominent potential candidate genes encoding for different categories of proteins.

4.3.1. Transporters

Protein synthesis during the grain-filling phase needs abundant supply of amino acids which are remobilized from the vegetative tissues. The transport of amino acids across membranes and translocation from source to sink is mediated by amino acid transporter proteins. Here, we found two candidate genes in chromosome 3A (*TraesCS3A02G484600* and *TraesCS3A02G484900*) whose product is an amino acid permease (Tables 2 and S5), a transmembrane protein from the Amino Acid/Auxin Permease family (AAP). Amino acid transporters are shown to be highly expressed in wheat grain cells and regulate the accumulation of amino acids in grain [54]. Recently, Jin et al. [55] reported on the cloning of three homoeologues of *TaAAP6* (on 3A, 3B and 3D) from *OsAAP6* (a putative amino acid permease gene) by homology cloning, and showed that *TaAAP6-3B* was a regulator of GPC, and its favored allele *TaAAP6-3B-I* has clear potential for application in wheat breeding for enhanced GPC. Based on recently published results, transporters including transporter proteins are expected to have a potential role in grain micronutrient accumulation [9,19,23].

4.3.2. Proteins, Implemented in Transcriptional and Post-Translational Regulation

The output of our candidate gene search included three genes in chromosome 3A conferring NAC domain-containing proteins (Tables 2 and S5). Recent transcriptomic analyses in wheat grains reported the expression of several transcription factors, including members of the NAM/ATAF/CUC (NAC) family [56,57]. The wild allele of a NAC gene (NAM-B1, NO APICAL MERISTEM) was reported to encode a transcription factor that accelerates senescence and increases nutrient remobilization from leaves to developing grains [20]. Moreover, the authors reported on positional cloning of *Gpc-B1* gene (NAM-B1), which significantly increases the levels of GPC, Fe and Zn. NAM genes have been recently reported to regulate the transcriptional changes during early leaf senescence in wheat including genes associated with nitrogen transport [58]. A GPC-associated SNP on chromosome 5D was found to encode a NAC domain superfamily protein (gene *TraesCS5D02G537600*) as well [50]. A recent study reported an endosperm-specific transcription factor TaNAC019 that regulates the accumulation of grain prolamins (glutenin) and starch, providing a candidate gene for breeding of wheat with improved grain quality [59]. Putative candidate genes for NAC transcription factors and transmembrane proteins were found to play a critical role in the Fe accumulation in wheat grains [8].

In our study, we found three post-translational putative genes in chromosome 3A with GO-identified biological process “protein phosphorylation”, molecular function “protein kinase activity” and functional annotation serine/threonine protein kinase or receptor-like kinase (Tables 2 and S5). Genes annotated as protein kinases have shown their potential role in macro- and micronutrients accumulation in wheat grain, as well [23]. One putative gene, *TraesCS3A02G488100*, detected on chromosome 3A, was associated with the biological process “protein glycosylation” and molecular function “acetylglucosaminyltransferase activity” encoding for a glycosyl transferase. Candidate genes coding for proteins with acetylglucosaminyltransferase or kinase activity have been predicted in *Aegilops tauschii* (the D-genome progenitor of bread wheat) genomic regions harboring grain micronutrient (Fe and Mn) MTAs [33].

Our candidate gene search identified two putative genes involved in another post-translational modification—protein ubiquitination: one gene for E3 ubiquitin-protein ligase and another one for the E2 ubiquitin-conjugating enzyme, both on chromosome 3A (Tables 2 and S5). We also identified six putative genes encoding F-box proteins on chromosomes 3A (5) and 3B (1) (Tables 2 and S5). This is in accordance with the reported high-confidence candidate gene (*TraesCS1B02G046800*) associated with GPC in wheat governing a F-box-like domain superfamily protein on chromosome 1B [52]. The F-box domain-containing proteins are key players in the ubiquitin-mediated protein turnover. Transcriptome analysis during various wheat developmental stages revealed that some F-box genes were specifically expressed at the seed developmental stages [60]. Interestingly, the putative candidate gene *TraesCS3A02G389000* linked to the strongest SNP marker (AX-95002032) on chromosome 3A associated with grain Fe concentration in wheat also was found to encode a F-box protein [19]. This observation hints at a possible link between protein and mineral homeostasis.

The stable significant SNP (*BS00062734_51*) on chromosome 3B identified in this study is located within a gene (*TraesCS3B02G339000*) that encodes for a plastidial lipoyl synthase (Tables 2 and S5), a protein responsible for the biosynthesis of lipoic acid and further involved in a rare but highly conserved lysine post-translational modification, known as protein lipoylation [61]. Interestingly, *BS00062734_51* was previously reported as the most significant marker in a QTL region on 3B for grain yield and duration of pre-anthesis sub-phases in a set of 210 winter genotypes selected from the GABI_WHEAT population [62].

4.3.3. Proteins, Involved in the Biosynthesis of Other Storage Materials

The biosynthesis of carbohydrates (mostly starch) is one of the most important metabolic processes in developing wheat grains besides the synthesis and accumulation of storage proteins. Wheat reserve starch is synthesized by the enzymatic machinery in the endosperm cells. Key proteins in the starch biosynthesis pathway have been recently identified [63]. In the current work, we identified a candidate gene on chromosome 3A (*TraesCS3A02G289800*), encoding PROTEIN TARGETING TO STARCH (PTST) (Tables 2 and S5)—a product that is among the 26 essential proteins in a starch biosynthesis model for wheat endosperm [63]. The suggested functional role of the wheat protein TaPTST1, an *Arabidopsis thaliana* AtPTST1 homolog, in starch metabolism targets the Granule-Bound Starch Synthase I (GBSSI) to starch granules in wheat endosperm [64]. Major starch-biosynthesis-related genes were shown to maintain high expression levels during the grain-filling stage in compliance with the high rates of starch accumulation [57,63,64]. Interestingly, the here-suggested candidate gene *TraesCS3A02G289800* for PTST is located in the vicinity of the significant QTN (*Excalibur_c13709_2568*) with a positive additive effect on GPC (0.54%, Table 1). As a major storage component of wheat endosperm, grain starch content (GSC) correlates positively with grain size and grain yield but shows a negative correlation with GPC. Interestingly, a recent study reported a genomic region on chromosome 6A that controlled both GPC and GSC with opposite allelic effects [46]. Our finding of a putative gene with a positive effect on GPC in a significant SNP-marked genomic region on chromosome 3A, that is involved in starch biosynthesis is suggestive

of probable genetic factors that escape the negative GPC-GSC correlation and, therefore, merits further attention.

In the present work, we detected a candidate gene (*TraesCS3A02G289900*) on chromosome 3A hit directly by the significant and consistent SNP (*Excalibur_c13709_2568*), and another one on chromosome 3B (*TraesCS3B02G337100*) (Tables 2 and S5) that annotates a metabolic protein–cellulose synthase. This enzyme is active in the plasma membrane–cell wall interface and is involved in the cell wall organization by controlling cellulose biosynthesis. It is also supposedly involved in the defense against pathogens [65]. A recent genome-wide analysis of the cellulose synthase-like (*Csl*) gene family in bread wheat revealed several *Csl* genes, especially concentrated on chromosomes of homoeologous groups 2 and 3 [66].

We found another candidate gene on chromosome 3A (*TraesCS3A02G484800*) hit directly by a significant and stable marker (*wsnp_Ku_c7811_13387117*; Table 1), which encodes for acetyl-CoA-carboxylase (Tables 2 and S5). Acetyl-CoA carboxylase (ACCase; EC 6.4.1.2) catalyzes the first committed step in de novo fatty acid biosynthesis. Earlier, the enzyme was shown to be present in both wheat germ and total wheat leaf protein [67].

4.3.4. Protective Proteins

During seed maturation, a series of mechanisms are induced in response to pathogens and environmental stress, in particular desiccation and elevated temperatures. Recent transcriptome and proteome research on wheat developing grain identified several important protein families with protective and detoxification activities [68]. In our study, numerous putative genes were found encoding beta-1,3-glucan hydrolases (beta-glucanases) on chromosome 3A, three genes for glutathione-S-transferases, also on 3A, as well as five genes encoding peroxidases, and two genes for lectin receptor kinases (LRKs) on chromosome 3B (Tables 2 and S5). Beta-glucanase expression in seeds plays a key role in the defense against seed pathogens and in the regulation of seed germination and dormancy [69]. Plant glutathione-S-transferases are another class of essential enzymes with detoxification functions during seed development [70], whereas peroxidases play many roles in plant seeds, including the elimination of toxic H₂O₂. The LRKs belong to a larger class of receptor-like kinases that play a vital role in the stimulation of immune responses in plants. In wheat, some genes for LRKs (*L-LRK7* and *L-LRK8*) were found to have relatively higher expression in seeds, being extremely upregulated under heat and drought stress [71].

A putative candidate gene *TraesCS3B02G339100* on chromosome 3B that is possibly related to GPC encodes for a subtilisin-like protease (Tables 2 and S5). Plant subtilisin-like proteins are involved in both disease resistance and immune priming [72]. Earlier, members of protease families, including subtilisin-like serine proteases were shown to participate in the N remobilization during grain filling [31], and in the breakdown and mobilization of reserve proteins from seeds to cotyledons during germination in wheat [73]. The recently reported functional annotation of the potential candidate genes for GPC in the 6A-chromosome physical region also indicated the presence of subtilisin-like proteases [46].

4.3.5. Structural Proteins

The presence of histones and ribosomal proteins was expected, as they are involved in the nucleosome assembly and ribosomes biogenesis, respectively. Our candidate gene search delivered one putative gene encoding histones on chromosome 3A, and one ribosomal protein encoding gene on chromosome 3B (Tables 2 and S5). Recently, Daba et al. [74] reported decreasing abundance of proteins involved in chromatin and ribosomal dynamics during grain development. On the contrary, Bonnot et al. [75] observed high demand for ribosome biogenesis proteins during the grain-filling and early maturation phases. We also identified one putative expansin-coding gene on chromosome 3A (Tables 2 and S5). Expansins are important proteins for cell wall loosening, cell enlargement, and extension of grain tissue, and play a vital role in determining grain size [76–78]. Recently, three

expansin-related genes that were mainly expressed during wheat kernel development were detected on chromosomes 1D, 4B, and 4D [74].

4.3.6. Metal Accumulation and Homeostasis

Based on our investigation, we found six potential genes in chromosomes 3A and 3B that are putatively related to Ca transport, i.e., genes encoding for Ca-channels and Ca-binding proteins (Tables 2 and S5). Putative candidate genes for Ca accumulation in seeds, including Ca-channels in addition to other genes that are possibly related to Ca transport, have been recently found by GWAS in a European wheat diversity panel of 353 accessions [7]. A very recent study reported on a histidine-rich Ca-binding protein on chromosome 3B (*TraesCS3B02G019900*) that may be closely related to Ca accumulation in wheat grain [79]. In addition, we detected one potential gene encoding for a metallothionein on chromosome 3A (Tables 2 and S5). Metallothioneins are cysteine-rich metal-binding proteins known to coordinate various transition metal ions including Zn, Cd and Cu thus maintaining metal homeostasis and detoxification [80]. These findings suggest that protein accumulation and metal ion transport and homeostasis in wheat grains might share common genetic mechanisms.

5. Conclusions

This study showed the importance of improving the grain protein content (GPC) in wheat by GWAS. A wide variation for GPC was found within the studied worldwide panel of 255 bread wheat accessions, along with significant genotypic and environmental effects with high broad-sense heritability estimates. The GWAS analysis conducted on the average BLUE values for GPC from three consecutive crop seasons identified seven significant marker-trait associations (MTAs) on chromosomes 1D, 3A (two MTAs), 3B, 3D, 4B and 5A. The loci on 3A and 3B are stable and novel ones underlying GPC in wheat. The most relevant putative candidate genes in the genomic regions on 3A and 3B included sequences encoding: (1) a subtilisin-like serine protease, with possible participation in N remobilization during grain filling; (2) amino acid permeases, having transporter functions; (3) transcription factors, protein kinases, and ubiquitin-related proteins possibly involved in transcriptional and post-translational regulatory functions; (4) enzymes from the biosynthetic pathways of starch, cellulose, and fatty acids; (5) protective proteins, including beta-gluconases, glutathione-S-transferases, peroxidases, and lectin receptor kinases; (6) histones, ribosomal proteins, and expansins, involved in chromatin, ribosomes, and cell wall dynamics during grain development; (7) proteins related to metal ions transport and homeostasis. Given the complexity of the GPC, many potential genes can be associated with the trait; therefore, this study provided a reference for genetic improvement of efficient wheat grain proteins for human nutrition and quality purposes. Moreover, the identification of nutritional genomic regions with pleiotropic effects on GPC and macro- and microminerals provides guidance for pyramid breeding of cultivars with higher nutritious value.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/foods12071399/s1>, Table S1: Information about a panel of 255 wheat accessions from 27 countries and summary statistics for GPC (%) from three growing seasons (2016, 2017, 2018), the average best linear unbiased estimates (BLUEs), and clusters; Table S2: Variability and stability of GPC (%) in a panel of 255 wheat accessions. Deviation from the annual average GPC from three growing seasons (2016, 2017, 2018); Table S3: Two-factor and one-factor analysis of variance; Table S4: Summary statistics for GPC (%) for a panel of 255 wheat accessions from three growing seasons (2016, 2017, 2018) and the average best linear unbiased estimates (BLUEs). Distribution by countries; Table S5: Candidate genes in the vicinity (± 2 Mbp) of the stable significant QTNs (MTAs); Figure S1: Monthly weather statistics for Gatersleben, Germany (51°49' N, 11°16' E) for three growing seasons. The temperature (A) and precipitation (B) are related to the climate norms (1961–90); Figure S2: The phylogenetic tree of 255 bread wheat cultivars constructed with the UPGMA algorithm based on 17,093 polymorphic SNPs from the 90 K SNP iSELECT array. Roman letters mark the suggested clusters.

Author Contributions: Conceptualization, S.M.; formal analysis, A.M.A., V.A. and M.A.R.A.; investigation, T.K., S.M., V.A. and D.D.; resources, A.B.; writing—original draft preparation, S.M. and A.M.A.; writing—review and editing, D.Z.A. and A.B.; visualization, A.M.A., V.A., S.M. and T.K.; supervision, S.M.; 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 Bulgarian National Science Fund, grant number KP-06-N31/17.

Data Availability Statement: Seeds from the bread wheat (*Triticum aestivum* L.) cultivars and lines were kindly provided by the Seed Genebank, Leibniz Institute for Plant Genetics and Crop Plants Research (IPK, Gatersleben), Germany.

Acknowledgments: The technical support of Katelina Prokopova is acknowledged.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the 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. IDRC. *Facts and Figures on Food and Biodiversity*; IDRC Communications, International Development Research Centre: Ottawa, ON, Canada, 2010; Available online: <https://www.idrc.ca/en/research-in-action/facts-figures-food-and-biodiversity> (accessed on 20 May 2022).
2. 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]
3. Acevedo, M.; Zurn, J.D.; Molero, G.; Singh, P.; He, X.; Aoun, M.; Juliana, P.; Bockleman, H.; Bonman, M.; El-Sohl, M.; et al. The role of wheat in global food security. In *Agricultural Development and Sustainable Intensification*, 1st ed.; Nagothu, U.S., Ed.; Taylor & Francis Group: London, UK, 2018; Chapter 4, pp. 81–110, ISBN 9780203733301. [CrossRef]
4. Obeidat, B.A.; Abdul-Hussain, S.S.; Al Omari, D.Z. Effect of addition of germinated lupin flour on the physiochemical and organoleptic properties of cookies. *J. Food Process. Preserv.* **2013**, *37*, 637–643. [CrossRef]
5. Tolve, R.; Simonato, B.; Rainero, G.; Bianchi, F.; Rizzi, C.; Cervini, M.; Giuberti, G. Wheat bread fortification by grape pomace powder: Nutritional, technological, antioxidant, and sensory properties. *Foods* **2021**, *10*, 75. [CrossRef] [PubMed]
6. Alomari, D.Z.; Eggert, K.; von Wirén, N.; Alqudah, A.M.; Polley, A.; Plieske, J.; Ganal, M.W.; Pillen, K.; Röder, M.S. Identifying candidate genes for enhancing grain Zn concentration in wheat. *Front. Plant Sci.* **2018**, *9*, 1313. [CrossRef]
7. 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. *Front. Plant Sci.* **2017**, *8*, 1797. [CrossRef]
8. Alomari, D.Z.; Eggert, K.; von Wirén, N.; Polley, A.; Plieske, J.; Ganal, M.W.; Liu, F.; Pillen, K.; Röder, M.S. Whole-Genome Association mapping and genomic prediction for iron concentration in wheat grains. *Int. J. Mol. Sci.* **2019**, *20*, 76. [CrossRef]
9. Gupta, P.K.; Balyan, H.S.; Sharma, S.; Kumar, R. Biofortification and bioavailability of Zn, Fe and Se in wheat: Present status and future prospects. *Theor. Appl. Genet.* **2021**, *134*, 1–35. [CrossRef] [PubMed]
10. Žilić, S.; Barać, M.; Pešić, M.; Dodig, D.; Ignjatović-Micić, D. Characterization of proteins from grain of different bread and durum wheat genotypes. *Int. J. Mol. Sci.* **2011**, *12*, 5878–5894. [CrossRef]
11. Shewry, P.R.; Halford, N.G. Cereal seed storage proteins: Structures, properties and role in grain utilization. *J. Exp. Bot.* **2002**, *53*, 947–958. [CrossRef]
12. Kumar, A.; Jain, S.; Elias, E.M.; Ibrahim, M.; Sharma, L.K. An overview of QTL identification and marker-assisted selection for grain protein content in wheat. In *Eco-Friendly Agro-Biological Techniques for Enhancing Crop Productivity*; Sengar, R., Singh, A., Eds.; Springer: Singapore, 2018; Chapter 11, pp. 245–274. [CrossRef]
13. Prasad, M.; Kumar, N.; Kulwal, P.L.; Röder, M.S.; Balyan, H.S.; Dhaliwal, H.S.; Gupta, P.K. QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theor. Appl. Genet.* **2003**, *106*, 659–667. [CrossRef]
14. Ruan, Y.; Yu, B.; Knox, R.E.; Zhang, W.; Singh, A.K.; Cuthbert, R.; Fobert, P.; DePauw, R.; Berraies, S.; Sharpe, A.; et al. Conditional mapping identified quantitative trait loci for grain protein concentration expressing independently of grain yield in Canadian durum wheat. *Front. Plant Sci.* **2021**, *12*, 642955. [CrossRef]
15. 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]
16. 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]
17. 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]

18. 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. [\[CrossRef\]](#)
19. Rathana, 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. [\[CrossRef\]](#)
20. Uauy, C.; Distelfeld, A.; Fahima, T.; Blechl, A.; Dubkovsky, J.A. NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* **2006**, *314*, 1298–1301. [\[CrossRef\]](#) [\[PubMed\]](#)
21. 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 3 wild emmer wheat RIL population. *Theor. Appl. Genet.* **2009**, *119*, 353–369. [\[CrossRef\]](#) [\[PubMed\]](#)
22. 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 Res.* **2012**, *138*, 57–62. [\[CrossRef\]](#)
23. Alomari, D.Z.; Alqudah, A.M.; Pillen, K.; von Wirén, N.; Röder, M.S. Toward identification of a putative candidate gene for nutrient mineral accumulation in wheat grains for human nutrition purposes. *J. Exp. Bot.* **2021**, *72*, 6305–6318. [\[CrossRef\]](#)
24. Krishnappa, G.; Rathana, N.D.; Sehgal, D.; Ahlawat, A.K.; Singh, S.K.; Singh, S.K.; Shukla, R.B.; Jaiswal, J.P.; Solanki, I.S.; Singh, G.P.; et al. Identification of novel genomic regions for biofortification traits using an SNP marker-enriched linkage map in wheat (*Triticum aestivum* L.). *Front. Nutr.* **2021**, *8*, 669444. [\[CrossRef\]](#)
25. 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.
26. STATISTICA, version 14; StatSoft Inc.: Tulsa, OK, USA, 2020.
27. Chen, W.C. Phylogenetic Clustering with R package Phyclust. Phyloclustering-Phylogenetic Clustering Website. 2010. Available online: <https://CRAN.R-project.org/package=phyclust> (accessed on 18 March 2023).
28. Sokal, R.R.; Michener, C.D. A Statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **1958**, *28*, 1409–1438.
29. Schierenbeck, M.; Alqudah, A.M.; Ulrike 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. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Roberts, I.N.; Caputo, C.; Kade, M.; Victoria Criado, M.; Barneix, A.J. Subtilisin-like serine proteases involved in N remobilization during grain filling in wheat. *Acta Physiol. Plant.* **2011**, *33*, 1997–2001. [\[CrossRef\]](#)
32. Jones, C.; Olson-Rutz, K. *Practices to Increase Wheat Grain Protein*; Montana State University Extension: Bozeman, MT, USA, 2020; bulletin EB0206.
33. Arora, S.; Cheema, J.; Poland, J.; Uauy, C.; Chhuneja, P. Genome-wide association mapping of grain micronutrients concentration in *Aegilops tauschii*. *Front. Plant Sci.* **2019**, *10*, 54. [\[CrossRef\]](#)
34. 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. [\[CrossRef\]](#)
35. Litvinenko, N.A. Grain quality indices in the program of winter bread wheat breeding for the steppe region of Ukraine. In *Wheat in a Global Environment*; Bedö, Z., Lang, L., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; pp. 273–278.
36. Hristov, N.; Mladenov, N.; Djuric, V.; Kondic-Spika, A.; Marjanovic-Jeromela, A. Improvement of wheat quality in cultivars released in Serbia during the 20th century. *Cereal Res. Comm.* **2010**, *38*, 111–121. [\[CrossRef\]](#)
37. Mladenov, V.; Banjac, B.; Krishna, A.; Milošević, M. Relation of grain protein content and some agronomic traits in European cultivars of winter wheat. *Cereal Res. Comm.* **2011**, *40*, 532–541. [\[CrossRef\]](#)
38. Khlestkina, E.K.; Pshenichnikova, T.A.; Usenko, N.I.; Otmakhova, Y.S. Prospective applications of molecular genetic approaches to control technological properties of wheat grain in the context of the “grain–flour–bread” chain. *Vavilov J. Genet. Breed.* **2016**, *20*, 511–527, (In Russian, English Summary). [\[CrossRef\]](#)
39. Abugalieva, A.; Peña, R.J. Grain quality of spring and winter wheat in Kazakhstan. *Asian Australas. J. Plant Sci. Biotechnol.* **2010**, *4*, 87–90.
40. Nazarenko, M.; Mykolenko, S.; Okhmat, P. Variation in grain productivity and quality of modern winter wheat varieties in northern Ukrainian Steppe. *Ukr. J. Ecol.* **2020**, *10*, 102–108. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Riaz, Q. Changes in Australian Wheat Quality over the Last 150 Years. Ph.D. Thesis, Charles Sturt University, Bathurst, Australia, 2020.
42. 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. [\[CrossRef\]](#)
43. Bordes, J.; Ravel, C.; Le Gouis, J.; Lapiere, A.; Charmet, G.; Balfourier, F. Use of a global wheat core collection for association analysis of flour and dough quality traits. *J. Cereal Sci.* **2011**, *54*, 137–147. [\[CrossRef\]](#)
44. Battenfield, S.D. Genomic Selection and Association Mapping for Wheat Processing and End-Use Quality. Ph.D. Thesis, Kansas State University, Manhattan, KS, USA, 2015.
45. Mir Dirkvand, R.; Najafan, G.; Bihamta, M.R.; Ebrahimi, A. Mapping some seed quality traits in bread wheat (*Triticum aestivum* L.) by association mapping using SSR markers. *J. Appl. Biotechnol. Rep.* **2018**, *5*, 92–99. [\[CrossRef\]](#)

46. Muqaddasi, Q.H.; Brassac, J.; Ebmeyer, E.; Kollers, S.; Korzun, V.; Argillier, O.; Stiewe, G.; Plieske, J.; Ganai, 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. [\[CrossRef\]](#)
47. Lou, H.; Zhang, R.; Liu, Y.; Guo, D.; Zhai, S.; Chen, A.; Zhang, Y.; Xie, C.; You, M.; Peng, H.; et al. Genome-wide association study of six quality-related traits in common wheat (*Triticum aestivum* L.) under two sowing conditions. *Theor. Appl. Genet.* **2021**, *134*, 399–418. [\[CrossRef\]](#)
48. 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. [\[CrossRef\]](#)
49. Govta, N.; Poldá, I.; Sela, H.; Cohen, Y.; Beckles, D.M.; Korol, A.B.; Fahima, T.; Saranga, Y.; Krugman, T. Genome-Wide Association Study in bread wheat identifies genomic regions associated with grain yield and quality under contrasting water availability. *Int. J. Mol. Sci.* **2022**, *23*, 10575. [\[CrossRef\]](#)
50. 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. [\[CrossRef\]](#)
51. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
52. 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. [\[CrossRef\]](#)
53. Jiang, G.-L. Molecular markers and marker-assisted breeding in plants. In *Plant Breeding from Laboratories to Fields*; Andersen, S.B., Ed.; InTech: London, UK, 2013; Volume 3, pp. 45–83. [\[CrossRef\]](#)
54. Wan, Y.; Wang, Y.; Shi, Z.; Rentsch, D.; Ward, J.L.; Hassall, K.; Sparks, C.A.; Huttly, A.K.; Buchner, P.; Powers, S.; et al. Wheat amino acid transporters highly expressed in grain cells regulate amino acid accumulation in grain. *PLoS ONE* **2021**, *16*, e0246763. [\[CrossRef\]](#)
55. Jin, X.; Feng, B.; Xu, Z.; Fan, X.; Lin, J.; Liu, Q.; Zhu, P.; Wang, T. *TaAAP6-3B*, a regulator of grain protein content selected during wheat improvement. *BMC Plant Biol.* **2018**, *18*, 71. [\[CrossRef\]](#)
56. Ma, D.; Gao, H.; Du, C.; Li, L.; Sun, W.; Liu, S.; Wang, C.; Xie, Y.; Kang, G. Transcriptomic and Metabolomics Analysis of Different Endosperm Region under Nitrogen Treatments. *Int. J. Mol. Sci.* **2019**, *20*, 4212. [\[CrossRef\]](#)
57. 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. [\[CrossRef\]](#)
58. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Gao, Y.; An, K.; Guo, W.; Chen, Y.; Zhang, R.; Zhang, X.; Chang, S.; Rossi, V.; Jin, F.; Cao, X.; et al. The endosperm-specific transcription factor TaNAC019 regulates glutenin and starch accumulation and its elite allele improves wheat grain quality. *Plant Cell* **2021**, *33*, 603–622. [\[CrossRef\]](#)
60. Hong, M.J.; Kim, J.-B.; Seo, Y.W.; Kim, D.Y. F-Box genes in the wheat genome and expression profiling in wheat at different developmental stages. *Genes* **2020**, *11*, 1154. [\[CrossRef\]](#)
61. Martínez-Noguerol, R.; Acket, S.; Troncoso-Ponce, M.A.; Garcés, R.; Thomasset, B.; Venegas-Calderón, M.; Salas Joaquín, J.; Martínez-Force, E.; Moreno-Pérez, A.J. Characterization of *Helianthus annuus* lipoic acid biosynthesis: The mitochondrial octanoyltransferase and lipoyl synthase enzyme system. *Front. Plant Sci.* **2021**, *12*, 781917. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Guo, Z.; Chen, D.; Röder, M.S.; Ganai, M.W.; Schnurbusch, T. Genetic dissection of pre-anthesis sub-phase durations during the reproductive spike development of wheat. *Plant J.* **2018**, *95*, 909–918. [\[CrossRef\]](#)
63. Gu, Y.; Han, S.; Chen, L.; Mu, J.; Duan, L.; Li, Y.; Yan, Y.; Li, X. Expression and regulation of genes involved in the reserve starch biosynthesis pathway in hexaploid wheat (*Triticum aestivum* L.). *Crop J.* **2021**, *9*, 440–455. [\[CrossRef\]](#)
64. Sharma, V.; Fandade, V.; Kumar, P.; Parveen, A.; Madhawan, A.; Bathla, M.; Mishra, A.; Sharma, H.; Rishi, V.; Satbhai, S.B.; et al. Protein targeting to starch 1, a functional protein of starch biosynthesis in wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* **2022**, *109*, 101–113. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Caño-Delgado, A.; Penfield, S.; Smith, C.; Catley, M.; Bevan, M. Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. *Plant J.* **2003**, *34*, 351–362. [\[CrossRef\]](#) [\[PubMed\]](#)
66. 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. [\[CrossRef\]](#)
67. Gornicki, P.; Haselkorn, R. Wheat acetyl-CoA carboxylase. *Plant Mol. Biol.* **1993**, *22*, 547–552. [\[CrossRef\]](#)
68. Yu, Y.; Zhu, D.; Ma, C.; Cao, H.; Wang, Y.; Xu, Y.; Zhang, W.; Yan, Y. Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *Crop J.* **2016**, *4*, 92–106. [\[CrossRef\]](#)
69. Leubner-Metzger, G. Functions and regulation of -1,3-glucanases during seed germination, dormancy release and after-ripening. *Seed Sci. Res.* **2003**, *13*, 17–34. [\[CrossRef\]](#)
70. Vaish, S.; Gupta, D.; Mehrotra, R.; Mehrotra, S.; Kumar Basantani, M. Glutathione S-transferase: A versatile protein family. *3 Biotech* **2020**, *10*, 321. [\[CrossRef\]](#) [\[PubMed\]](#)

71. Shumayla; Sharma, S.; Pandey, A.K.; Singh, K.; Upadhyay, S.K. Molecular characterization and global expression analysis of lectin receptor kinases in bread wheat (*Triticum aestivum*). *PLoS ONE* **2016**, *11*, e0153925. [[CrossRef](#)] [[PubMed](#)]
72. Figueiredo, A.; Monteiro, F.; Sebastiana, M. Subtilisin-like proteases in plant–pathogen recognition and immune priming: A perspective. *Front. Plant Sci.* **2014**, *5*, 793. [[CrossRef](#)]
73. 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](#)]
74. Daba, S.D.; Liu, X.; Aryal, U.; Mohammadi, M. A proteomic analysis of grain yield-related traits in wheat. *AoB PLANTS* **2020**, *12*, plaa042. [[CrossRef](#)]
75. Bonnot, T.; Bancel, E.; Chambon, C.; Boudet, J.; Branlard, G.; Martre, P. Changes in the nuclear proteome of developing wheat (*Triticum aestivum* L.) grain. *Front. Plant Sci.* **2015**, *6*, 905. [[CrossRef](#)] [[PubMed](#)]
76. Lizana, X.C.; Riegel, R.; Gomez, L.D.; Herrera, J.; Isla, A.; McQueen-Mason, S.J.; Calderini, D.F. Expansins expression is associated with grain size dynamics in wheat (*Triticum aestivum* L.). *J. Exp. Bot.* **2010**, *61*, 1147–1157. [[CrossRef](#)]
77. Marowa, P.; Ding, A.; Kong, Y. Expansins: Roles in plant growth and potential applications in crop improvement. *Plant Cell Rep.* **2016**, *35*, 949–965. [[CrossRef](#)]
78. Tillett, B.J.; Hale, C.O.; Martin, J.M.; Giroux, M.J. Genes impacting grain weight and number in wheat (*Triticum aestivum* L. ssp. *aestivum*). *Plants* **2022**, *11*, 1772. [[CrossRef](#)] [[PubMed](#)]
79. Shi, X.; Zhou, Z.; Li, W.; Qin, M.; Yang, P.; Hou, J.; Huang, F.; Lei, Z.; Wu, Z.; Wang, J. Genome-wide association study reveals the genetic architecture for calcium accumulation in grains of hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2022**, *22*, 229. [[CrossRef](#)]
80. Yuan, J.; Chen, D.; Ren, Y.; Zhang, X.; Zhao, X. Characteristic and expression analysis of a metallothionein gene, *OsMT2b*, down-regulated by cytokinin suggests functions in root development and seed embryo germination of rice. *Plant Physiol.* **2008**, *146*, 1637–1650. [[CrossRef](#)]

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.