



Article

Multomics Based Association Mapping in Wheat Reveals Genetic Architecture of Quality and Allergenic Related Proteins

Khaoula El Hassouni ¹, Muhammad Afzal ¹, Kim A. Steige ¹, Malte Sielaff ², Valentina Curella ³, Manjusha Neerukonda ³, Stefan Tenzer ², Detlef Schuppan ^{3,4}, Carl Friedrich Horst Longin ^{1,*} and Patrick Thorwarth ¹

- ¹ State Plant Breeding Institute, University of Hohenheim, Fruwirthstr. 21, 70599 Stuttgart, Germany; khaoula.elhassouni@uni-hohenheim.de (K.E.H.); m.afzal@uni-hohenheim.de (M.A.); kim.steige@uni-hohenheim.de (K.A.S.); patrick.thorwarth@uni-hohenheim.de (P.T.)
- ² Institute for Immunology, University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany; malte.sielaff@uni-mainz.de (M.S.); tenzer@uni-mainz.de (S.T.)
- ³ Institute of Translational Immunology and Research Center for Immune Therapy, University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany; cvalenti@uni-mainz.de (V.C.); manjusha@uni-mainz.de (M.N.); detlef.schuppan@unimedizin-mainz.de (D.S.)
- ⁴ Division of Gastroenterology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Boston, MA 02215, USA
- * Correspondence: friedrich.longin@uni-hohenheim.de

Abstract: Wheat is an important staple crop since its proteins contribute to human and animal nutrition and are important for its end-use quality. However, wheat proteins can also cause adverse human reactions for a large number of people. We performed a genome wide association study (GWAS) on 114 proteins quantified by LC-MS-based proteomics and expressed in an environmentally stable manner in 148 wheat cultivars with a heritability > 0.6. For 54 proteins, we detected quantitative trait loci (QTL) that exceeded the Bonferroni-corrected significance threshold and explained 17.3–84.5% of the genotypic variance. Proteins in the same family often clustered at a very close chromosomal position or the potential homeolog. Major QTLs were found for four well-known glutenin and gliadin subunits, and the QTL segregation pattern in the protein encoding the high molecular weight glutenin subunit *Dx5* could be confirmed by SDS gel-electrophoresis. For nine potential allergenic proteins, large QTLs could be identified, and their measured allele frequencies open the possibility to select for low protein abundance by markers as long as their relevance for human health has been conclusively demonstrated. A potential allergen was introduced in the beginning of 1980s that may be linked to the cluster of resistance genes introgressed on chromosome 2AS from *Triticum ventricosum*. The reported sequence information for the 54 major QTLs can be used to design efficient markers for future wheat breeding.

Keywords: wheat; proteins; quality; allergen; QTL; breeding



Citation: El Hassouni, K.; Afzal, M.; Steige, K.A.; Sielaff, M.; Curella, V.; Neerukonda, M.; Tenzer, S.; Schuppan, D.; Longin, C.F.H.; Thorwarth, P. Multomics Based Association Mapping in Wheat Reveals Genetic Architecture of Quality and Allergenic Related Proteins. *Int. J. Mol. Sci.* **2023**, *24*, 1485. <https://doi.org/10.3390/ijms24021485>

Academic Editor: Pedro Martínez-Gómez

Received: 21 December 2022

Revised: 10 January 2023

Accepted: 10 January 2023

Published: 12 January 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

Wheat (*Triticum aestivum* ssp. *aestivum*) is the most widely grown crop and a major component of the human diet worldwide. This staple crop is one of the most important sources of energy [1] and on average provides 20% of the total protein and calories in human nutrition [2]. Wheat is consumed in many different forms, and each type of end-product requires a particular quality based on the viscoelastic properties of the dough, which are mainly influenced by the amount and composition of gluten [3]. Gluten accounts for approximately 80% of the total protein in the grain and can be divided into gliadins and glutenins. Glutenins are classified into high and low molecular weight subunits (HMW-GS and LMW-GS) [4], which are encoded by the loci *Glu-1* and *Glu-3*, respectively. Differences between allele pairs in glutenin subunits have a strong influence on the end-use

quality [5]. For example, several studies have shown that the alleles *Dx5 + Dy10* (*Glu-D1d*) are associated with high quality, whereas *Dx2 + Dy12* (*Glu-D1a*) lead to poor quality [6–8]. Consequently, wheat breeders have intensively selected for specific combinations of HMW-GS since the pioneering work of Payne et al. [9]. In addition to its effect on end-product quality, gluten and non-gluten proteins, such as wheat amylase trypsin inhibitors (ATIs), are also associated with various human health disorders, such as celiac disease, allergic reactions, and non-celiac wheat sensitivity [10–12]. However, targeting wheat allergens has never been a goal of breeding programs other than improving gluten quality due to its impact on end-product quality. For instance, the abundance of many allergenic proteins or ATIs in wheat cultivars released in the last century has not changed [13,14]. While only a few proteins have been investigated in detail in recent decades [15], the recent developments in the field of mass spectrometry-based proteomics has led to the possibility to determine hundreds of proteins in a single sample [16]. Afzal et al. [17] analyzed the flour proteome of 15 spelt and wheat cultivars grown in three different locations and identified 3050 proteins, including 300 proteins with moderate-to-high heritability (>0.4). However, to our knowledge, no study investigated the genetic architecture of the large number of proteins that have been discovered with modern proteomic tools so far.

Therefore, we performed a GWAS to investigate the genetic architecture of 114 proteins quantified using liquid chromatography-mass spectrometry (LC-MS)-based label-free quantitative (LFQ) proteomics from 148 bread wheat cultivars grown in three environments. In addition, the genetic and temporal trend of the alleles of some major QTLs associated with relevant proteins were investigated, and the sequence information for relevant QTLs is provided so that it is possible to design molecular markers.

2. Results

In a previous study, we measured 756 proteins in aqueous extracts from the whole-grain flour of 148 wheat cultivars grown in three environments [18]. Only 114 of these 756 proteins had a stable expression across environments in most wheat cultivars with a heritability larger than 0.6. For these 114 proteins, we performed a GWAS and detected QTLs for 54 proteins that exceeded the Bonferroni-corrected significance threshold (Figures 1 and 2). For all these 54 proteins, a single major QTL explaining 17.3 to 84.5% of the genotypic variance was identified (Figure 2b, Table 1, Table S1). For 24 proteins, the identified QTLs explained >50% of the genotypic variance (Figure 2b). In contrast, for 60 proteins, no marker-trait association was detected that exceeded the Bonferroni-corrected significance threshold.

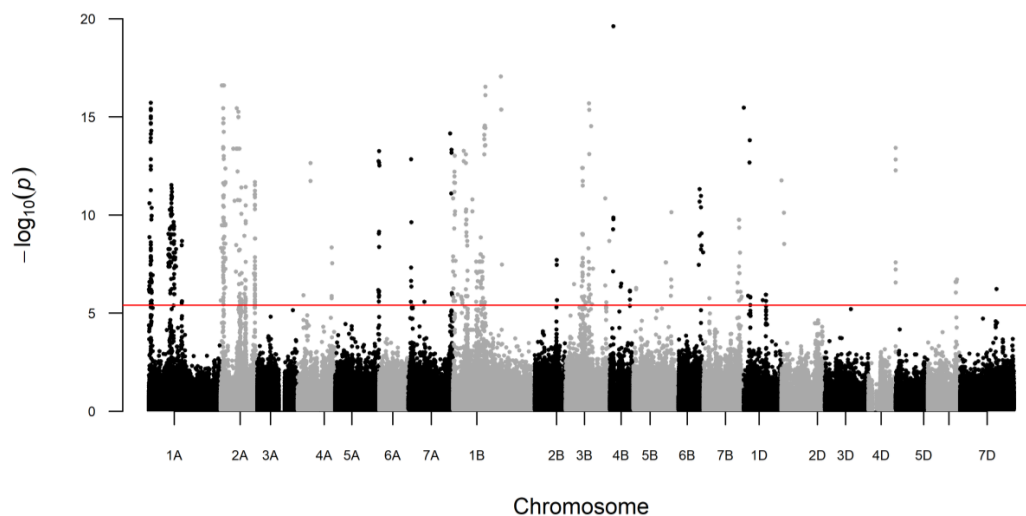


Figure 1. Manhattan plot showing significant marker-trait associations for all 54 proteins, where QTLs at Bonferroni-corrected significance threshold of $p < 0.05$ were identified (red line).

Table 1. QTLs controlling two gluten proteins, four non-gluten allergenic proteins and five gluten proteins, which were also listed as allergens in Allergome database; identified in this study (HMW = High molecular weight; LMW = Low molecular weight; LTP = Lipid transfer protein).

	Trait	UniProt ID	UniProt Name	Marker	Chr.	<i>p</i> -Value	LOD	Gen Pos. (cM)	Phy Pos. Start (bp)	Phy Pos. Stop (bp)	<i>P_G</i>	α -Effect
Gluten	prot051	A0A3B6I2R2		1212683D	4A	4.65×10^{-9}	8.33	2,310,811	725,135,595	725,135,664	45.53	7006.50
	prot104	A0A3B6RB62		1283144D	7A	1.48×10^{-13}	12.83	211,140	16,051,783	16,366,772	54.70	13,071.91
Gluten and allergen	prot017	G1E6K7	HMW- <i>Dx5</i>	999473S	1D	1.18×10^{-6}	5.93	1,492,346	414,170,821	414,170,890	42.73	−1947.51
	prot028	Q94G97	γ -Gliadin	3027396D	1A	1.93×10^{-16}	15.72	150,899	4,462,173	4,462,215	82.41	12,546.17
	prot066	D2KFH0	Gliadin/avenin	999766S	3B	2.07×10^{-16}	15.68	1,607,210	734,966,174	734,966,228	82.01	5516.69
	prot139	I1XB56	LMW-GS	1201279D	1B	9.60×10^{-14}	13.02	182,700	6,301,587	6,301,656	73.82	10,631.02
	prot203	C3VN75	LMW- <i>GluA3</i>	1043336S	1A	2.51×10^{-9}	8.60	149,830	1,235,911	1,235,980	49.85	−12,552.71
Non-gluten allergen	prot008	Q6W8Q2	Peroxiredoxin	1039232S	2A	5.61×10^{-16}	15.25	1,198,150	98,916,927	102,957,935	75.2	6937.00
	prot189	Q2PCC3	LTP	1203564D	5B	7.33×10^{-11}	10.13	2,583,420	690,570,279	690,570,348	32.32	3056.54
	prot235	Q2PCC5	LTP	1219351D	5A	5.60×10^{-14}	13.25	2,978,100	707,857,285	707,857,336	56.53	−47,278.73
	prot288	A0A1D5UB33		1126529D	2A	2.51×10^{-17}	16.60	80,770	5,395,135	10,230,260	84.51	−34,444.07

Chr. Chromosome, Gen Pos. chromosome position in cM, Phy Pos. start/stop sequence position in base pairs according to the bread wheat reference genome (IWGSC RefSeq v2.1), *P_G* proportion of genotypic variance explained by the QTL in percent, and allele substitution α -effect.

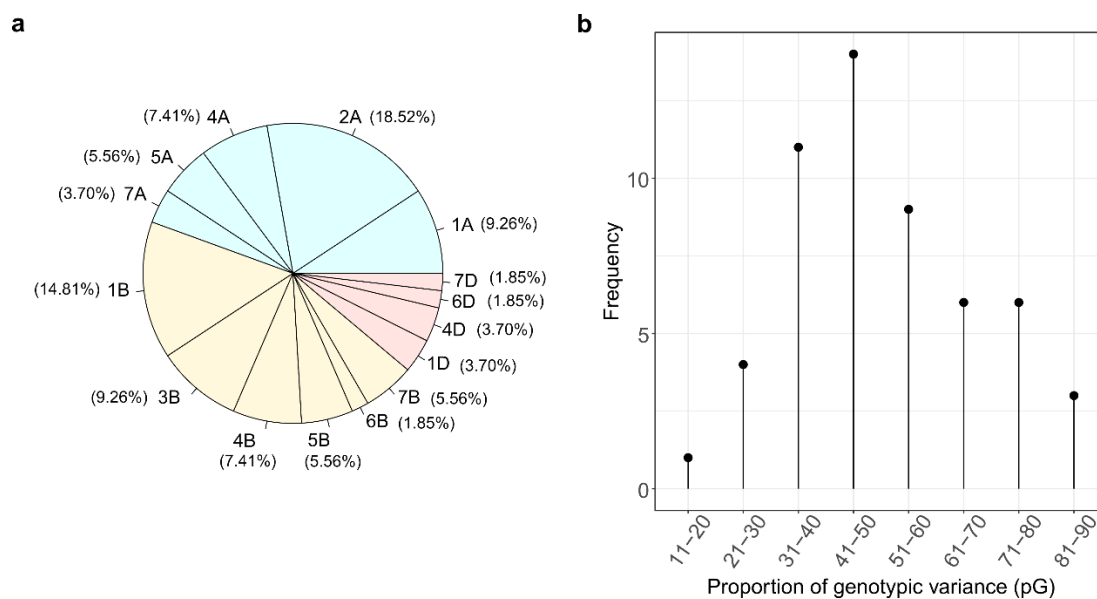


Figure 2. (a) Pie chart showing the distribution of the 54 major QTLs on different chromosomes; (b) frequency of the proportion of explained genotypic variance by the 54 QTLs.

The major QTLs were distributed across many chromosomes and partly clustered in similar chromosomal regions (Figures 2 and 3). QTLs were identified on chromosomes 1A, 2A, 4A, 5A, 7A, 1B, 3B, 4B, 5B, 6B, 7B, 1D, 4D, 6D, and 7D with a relatively similar distribution on the A and B genomes, whereas only 12% of the identified QTLs were located on the D genome. A higher number of proteins seemed to be affected by the major QTLs on chromosomes 1A, 2A, 1B, and 3B. Interestingly, some QTLs for different proteins were identified at almost the same genomic position (Figure 3, Table S1). For instance, on chromosome 5A, the QTLs for prot085 and prot141, both β -amylases according to the UniProt database, had the same chromosomal and physical position. Similarly, the QTLs for prot171 and prot179 had the same physical position on chromosome 3B, but according to the various protein annotation databases available, it is not yet clear whether these proteins belong to the same family (Table S1). To design easy-to-use markers for breeding, we have summarized the SNP, genomic position, and sequence information of all identified 54 QTLs in Table S1.

We further investigated the allele frequencies of QTLs associated with these eleven proteins (Figure 4). For prot008, prot017, prot066, and prot235, allele frequencies of their QTLs were found to be around 0.5. In contrast, for the QTLs of the other seven proteins, allele frequencies tended to a considerably higher frequency of one allele. Interestingly, for five of these seven proteins, the QTLs alleles increasing protein abundance were more frequent. As the wheat cultivars used in this study were released in different decades of the last century, we grouped them accordingly to visualize potential selection trends by wheat breeders. For four proteins, we observed shifts in QTLs allele frequencies across the decades of breeding.

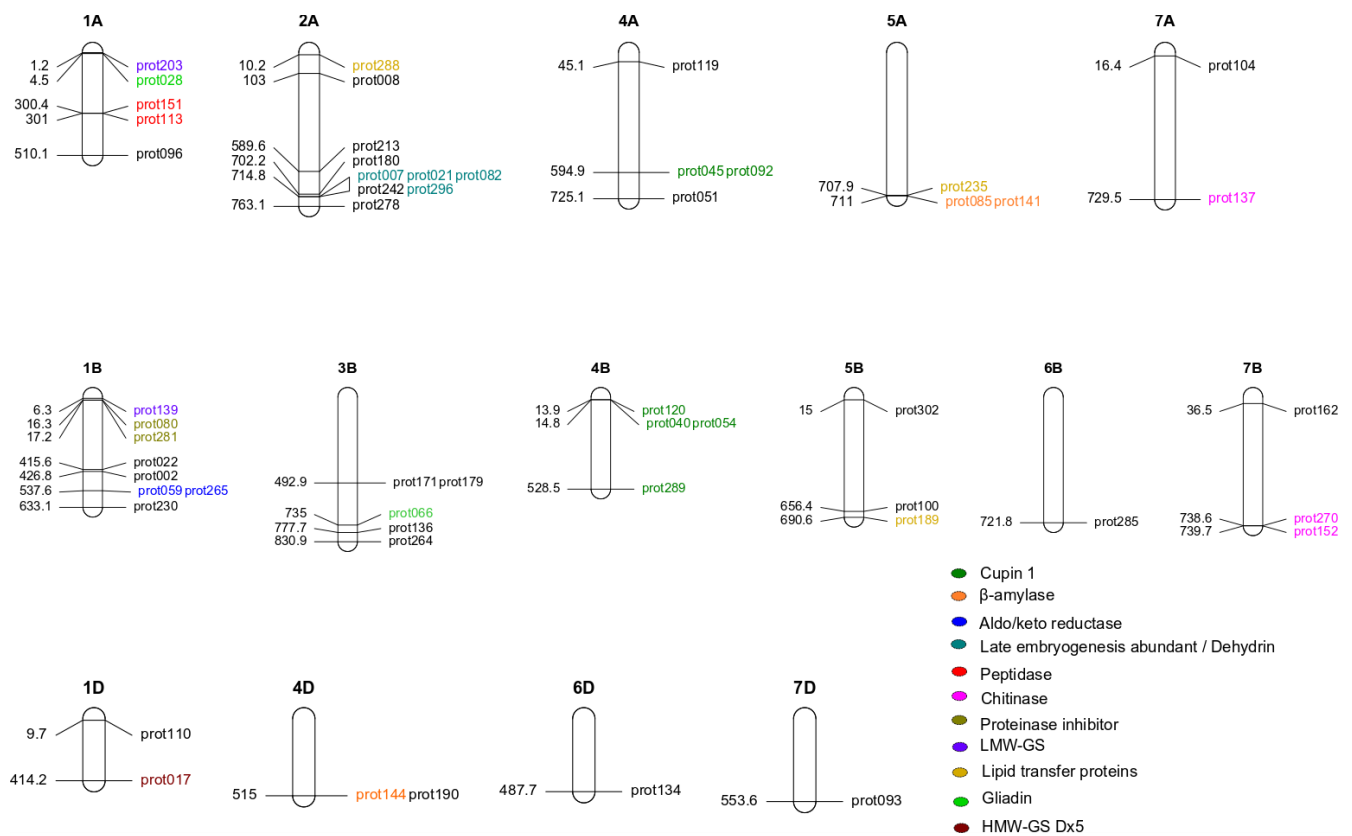


Figure 3. Chromosome map showing the distribution of 54 QTLs with their respective protein name and physical position (Mbp). For a detailed investigation of wheat flour proteins, we focused on the eleven proteins associated with gluten and potential allergenicity based on UniProt and InterPro databases (Table 1). Five of the seven gluten proteins were also present in the Allergome database (<https://allergome.org/>, accessed on 16 February 2021). We assigned these eleven proteins to the following groups: gluten, gluten and allergen, and non-gluten allergen. Using the UniProt database, we were able to name eight of these eleven proteins, including proteins important for baking quality such as HMW-GS *Dx5* and LMW *Glu-A3*. QTLs for two lipid transfer proteins could be identified, but not for other known, potentially allergenic wheat proteins such as ATIs or serpins.

Finally, we investigated the chromosomal regions harboring the QTLs of the 11 proteins in detail (Table S4). For these regions, we extracted high confidence (HC) genes from the bread wheat reference genome (IWGSC RefSeq v2.1) and evaluated these as potential candidate genes with functional annotations in the Pfam and InterPro databases similar to the different domains of gluten and allergenic proteins. Eight, five, and twenty-two potential candidate genes were identified in the QTLs target regions associated with gluten proteins (prot051 and prot104), allergens and gluten (prot017, prot028, prot139 and prot203), and non-gluten allergens (prot189, prot235), respectively (Table S4). No potential candidate genes could be identified for the QTLs detected for the prot066, prot008, and prot288.

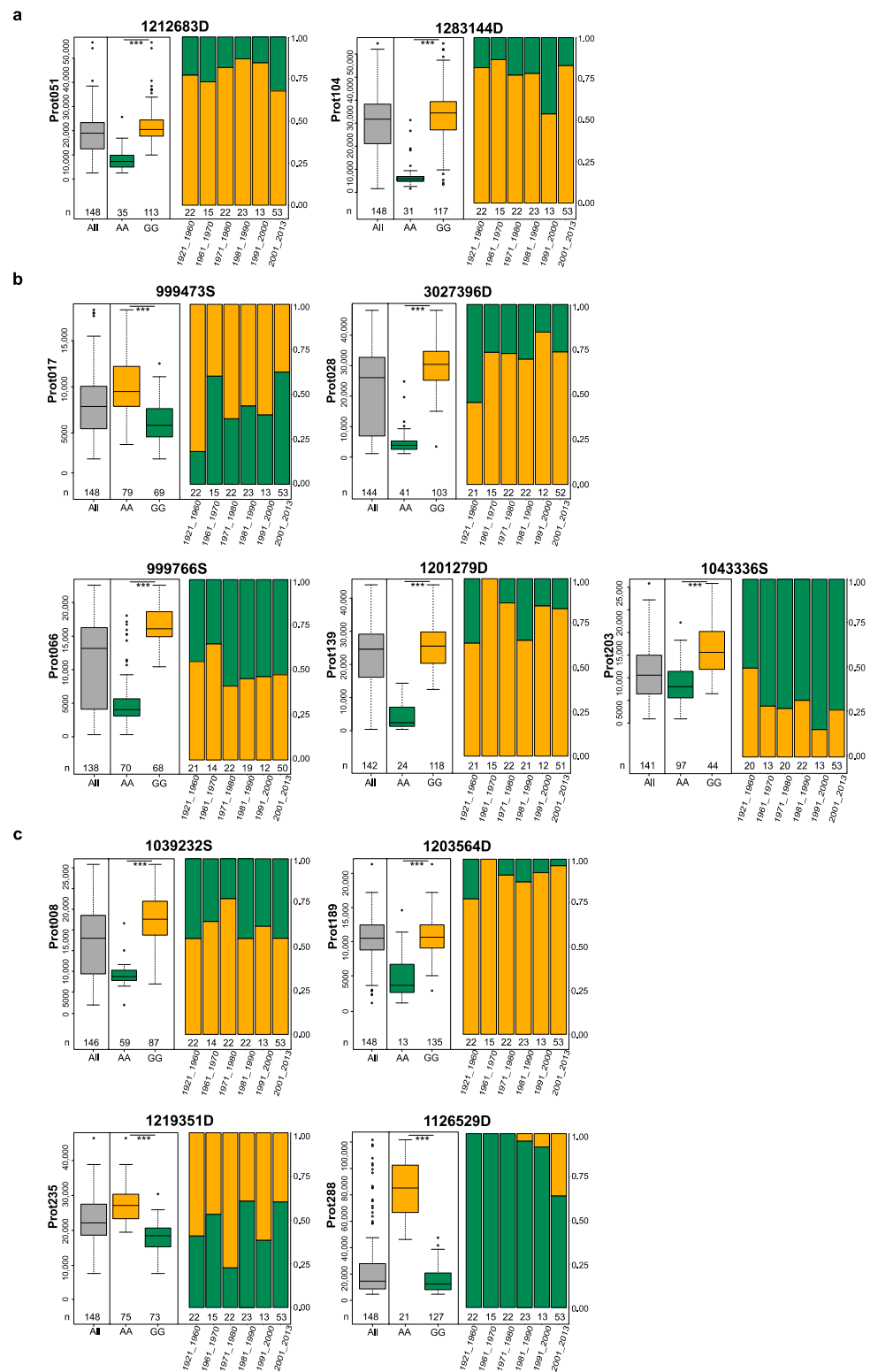


Figure 4. The effect of the major QTLs for gluten proteins prot051 and prot104 (a); gluten and allergenic proteins prot017, prot028, prot066, prot139, and prot203 (b); and non-gluten allergenic proteins prot008, prot189, prot235, and prot288 (c), and their allele frequencies according to the cultivar’s year of release (numbers below boxes represent the number of cultivars in the respective group). The leftmost boxplot in gray show the protein values for all cultivars. The protein-increasing allele is colored in orange; The protein-decreasing allele is colored in green. *** indicates significant difference at $p < 0.001$ between the two groups of cultivars containing contrasting alleles of a given marker.

3. Discussion

3.1. Major QTLs Identified for 54 Proteins

Implementing a GWAS using statistically conservative Bonferroni-corrected significance threshold, we identified major QTLs for 54 out of 114 proteins (Figures 1 and 2). Our findings suggest that more than half of the investigated proteins are quantitatively inherited and controlled by many genes, each with rather small effect. This quantitative inheritance is well-described in literature for the most investigated traits, e.g., yield, but also classically determined protein content [19–21]. In contrast, many of the QTLs identified for the 54 proteins had very high peaks in the Manhattan plot (Figure 1) and explained a large proportion of the genotypic variance of the individual proteins (Figure 2b, Table 1). In wheat, major QTLs are known, such as for plant height (*Rht* genes), heading time (*Ppd* genes), and disease resistance (e.g., *Lr* genes), but in most cases, the proportion of the explained genotypic variance was much lower than for many proteins in our study [22,23]. Therefore, the identified QTLs could be very interesting for future wheat breeding, provided that the relevance of the respective proteins for future wheat supply chains is demonstrated.

The 54 identified QTLs were similarly distributed across the A and B genomes, but only a small number of them were detected on the D genome (Figures 2a and 3). This is in line with the literature on genomics in wheat [24,25] and can be explained by the limited genetic diversity of the D genome compared to the A and B genomes. Interesting breeding approaches have begun to utilize the genetic potential of the D genome of wheat, such as synthetic wheat [26,27]. As these breeding lines are quite new and, to our knowledge, not yet present in European wheat cultivars, they were also not present in our wheat cultivar list.

For proteins belonging to the same family, we found that they are controlled by loci whose physical positions are located close to one another on the same chromosome or by loci on potential homologous chromosomes (Figure 3). For instance, we identified QTLs for six Cupin 1 proteins, all located on chromosomes 4A and 4B (Figure 3, Table 1). QTLs of proteins 045 and 092 were located on the identical physical map position on 4A, whereas QTLs for proteins 040, 054, and 120 were very close to each other on 4B. We found further QTLs clusters for other protein families. These QTLs are found for three late embryogenesis abundant proteins on 2A, two proteins of the aldo/keto reductase family on 1B, two β -amylases on 5A, and two proteins of chitinase class 1 on 7B. Potentially homologous chromosomal positions were identified for QTLs of two lipid transfer proteins on 5A and 5B, two plant antimicrobial proteins on 6B and 6D, three proteins of chitinase class 1 on 7A and 7B, and for two LMW-GS on 1A and 1B (Figure 3, Table 1). These findings are comparable to other traits where important gene families are located on the same group of homologous chromosomes, e.g., for plant height on chromosomes of group 4 (*Rht1* and *Rht2* genes) or heading time on chromosomes of group 2 (*Ppd-1* genes). In summary, to our knowledge, this largest GWAS study on the wheat proteome revealed a similar genetic architecture of proteins as reported for other traits, with major QTLs for 24 out of 114 proteins.

3.2. QTLs for Important Gluten Proteins

High and low molecular weight glutenins are of great importance for wheat end-use quality. They have been under intensive research and use in wheat breeding since the pioneering work of Payne and colleagues [4,9]. We identified three major QTLs underpinning three proteins specifically related to glutenins, one HMW- and two LMW-GS proteins (Table 1). On chromosome 1D, we found a major QTL explaining 42.7% of the genotypic variance for prot017, which is annotated as HMW-*Dx5* according to the Uniprot database (Table 1). Electrophoretic analysis by SDS-PAGE on our wheat cultivars revealed that the QTL allele GG of prot017 was represented by HMW banding unit 5 and the QTL allele AA by the HMW banding unit 2 for a total of 143 out of 148 wheat cultivars (Table S2), confirming the Uniprot annotation. Interestingly, the allelic difference from the SDS banding pattern could be deduced from the quantitative measurements of a single protein

(017), where cultivars with the *Dx5* (GG) unit had a lower abundance than cultivars with the *Dx2* (AA) unit (Figure 4, Table S2). While it was not possible to qualitatively distinguish the highly homologous protein isoforms by the tryptic peptides quantified by our mass-spectrometry-based proteomics workflow, the apparently allele-dependent expression level of the gene product was perfectly captured by the complementary SDS-PAGE approach. This case study highlights that quantitative protein measurements can provide additional relevant information content to purely genetic analyses for breeding studies.

In addition, we observed a temporal trend in selection for the QTL alleles of prot017. Wheat cultivars released after 2001 had a higher frequency of the *Dx5* allele (GG) than the *Dx2* allele (AA) compared with wheat cultivars bred before the year 2000 (Figure 4). Wheat breeders have intensively selected for the *Dx5* unit since the 1980s using SDS gel-electrophoresis, increasingly combined with molecular markers [28,29] such as the codominant markers 'UMN25' and 'UMN26' developed by Liu et al. [30] and the SNP markers used to discriminate Dx subunits developed and validated by Schwarz et al. [31]. However, >30% of modern wheat cultivars in our panel carry the allele *Dx2* (AA), showing further potential for improving baking quality. Since our panel does not contain any very modern wheat varieties, it is likely that breeders have already further increased the frequency of *Dx5*.

The major QTLs identified for two LMW proteins (prot139, prot203) were in similar regions on chromosome 1A and 1B (Figure 3). According to the UniProt database, prot203 corresponds to *Glu-A3* (Table S1), and we speculate that prot139 might be the homologous *Glu-B3* warranting further confirmation. For prot203 (= *Glu-A3*), a selection trend towards fixation of the QTL allele AA was evident over the different decades of wheat breeding (Figure 4). For prot139, the QTL allele that increases protein abundance appears to be close to fixation. The LMW glutenin subunits are much more difficult to identify than the HMW-GS described above due to their complexity, heterogeneity, and similarity to each other, as well as to some gliadins [32], and thus, have not been directly selected by wheat breeders in the last decades. The visible selection trend could come from indirect selection for these proteins by measuring dough and baking quality. Nevertheless, molecular markers capable of distinguishing sixteen different alleles at *Glu-A3* and *-B3* were recently developed [32], which will somewhat facilitate future selection.

Close by the genomic location of *Glu-A3*, we identified a major QTL explaining 82.4% of the genotypic variance of prot028, which is a γ -gliadin according to the UniProt database (Table 1). It is known that *Glu-A3*, *Glu-B3*, and *Glu-D3* are tightly linked to *Gli-A1*, *Gli-B1*, and *Gli-D1*, respectively, the latter representing multigene families encoding γ - and ω -gliadin subunits [33]. As with *Glu-A3*, a selection trend in wheat cultivars towards fixation of the protein-increasing QTL allele was evident for prot028 (Figure 4). Consequently, the selection for dough and baking quality in wheat has continuously modified the frequency of protein abundance of *Glu-A3* but also of γ -gliadin. In summary, new proteomic approaches were used to confirm the results of single protein analyses. Those approaches are now able to deliver thousands of proteins per sample [34], paving the way for much deeper exploration of their expression and relationship to agronomic and quality traits in crops.

3.3. Possibility to Breed for Low Allergen Content

Although wheat is an important and mostly healthy staple crop, a sizeable number of people suffer from wheat sensitivities, with most potential triggers being proteins [34]. We followed the approach of Zimmermann et al. [35] and Afzal et al. [34] and compiled a list of allergens based on data on seed-borne wheat allergens [36] and the Allergome database (<http://www.allergome.org/index.php> accessed on 16 February 2021) [37]. For nine proteins from this list, we detected major QTLs in our study (Table 1), which explained between 32.3% and 84.5% of the genotypic variance of the respective protein. Five of these were gluten proteins, two probable lipid transfer proteins, one peroxiredoxin, and one a potential protease inhibitor (Table S1). For three out of these nine proteins, we could observe a selection trend at the major QTLs in the wheat cultivars from the past

decades (Figure 4). The marker allele producing high protein abundance was increased for prot288, a protease inhibitor, whereas the marker allele responsible for low protein abundance was increased for prot017, HMW-*Dx5*, and prot203, LMW-GS *Glu-A3*. The latter two are important wheat-quality proteins that plant breeders have intensively selected for, as discussed earlier, and at the same time, appear to be potential allergens for a small number of people. Interestingly, the better baking quality at these two loci appears to be correlated with lower protein abundance, i.e., lower allergen levels.

The selection trend for prot288 is interesting in that the QTL allele, which increases protein abundance was introgressed in the early 1980s and its frequency then steadily increased by wheat breeders. Selection for or against potentially allergenic proteins has never been a goal in wheat breeding. Therefore, this selection trend may be due to the linkage with another target trait in wheat breeding that has been used since the 1980s and is largely influenced by the genomic region on the short arm of chromosome 2A. This chromosomal region contains an introgression from *Triticum ventricosum* that has a roughly comparable history [38]. This introgression carries several important disease-resistance genes (e.g., *Lr37* and *Sr38-Yr17-Lr34*) to various important rust diseases [39,40]. In addition, the introgression also appears to improve yield stability [41] and resistance to rice blast [42], all traits that are of great importance for many wheat breeding programs worldwide. For a large proportion of our wheat cultivars, molecular marker information for the disease resistance cluster *Sr38-Yr17-Lr34* (Table S3) is available, which matches almost perfectly with the different QTL alleles of prot288. Consequently, the increase in the QTL allele that increases the abundance of the potential allergen prot288 could be due to indirect selection of disease-resistance genes nearby. According to the physical positions, our identified QTL is 6 Mbp away from the locus reported for the disease resistance cluster. Future studies will have to show whether this potential linkage can be broken by targeted selection using markers for both loci.

For the QTLs of six potential allergenic proteins, we did not detect clear selection trends over the decades of wheat breeding, but either an almost fixation on the QTL allele causing high protein abundance (prot139, prot189) or similar frequencies of both alleles. This is confirmed by our companion study in which we quantified the absolute protein amounts of eight ATIs by isotopically labeled standard peptides [14]. Therein, major QTLs were identified for monomeric and dimeric ATIs with similar allele frequencies for the monomeric ATI 0.28 but near fixation of the QTL allele responsible for high protein abundance of the dimeric ATI 0.19-like. Consequently, the reported sequence information for the major QTLs identified in both studies could largely facilitate breeding for the low protein abundance of eleven potentially allergenic proteins. Further studies are, therefore, urgently needed to work out the relevance of reducing these proteins abundance for human and animal health, so that the laborious breeding progress for these additional traits can finally be addressed.

4. Materials and Methods

4.1. Plant Material and Field Experiments

Details of the plant material and field experiments were originally reported by Rapp et al. [43] Briefly, a panel of 148 bread wheat cultivars originating from different European countries and registered between 1921 and 2013 were grown in three different locations in Germany (Hohenheim, Oberer Lindenhof, and Eckartsweier). Field trials were executed using a partially replicated (P-rep) design with a net plot size of 1.25 m². The list of the cultivars and their details are provided in Table S5.

4.2. LFQ Proteomic

LFQ bottom-up proteomics data were retrieved from a previous study by Afzal et al. [18], publicly available as ProteomeXchange [44] dataset with the identifier PXD023654, in which the proteomics workflow was described in detail. Briefly, water/salt-soluble proteins were extracted from the full-kernel flours of the wheat samples described in Section 3.1

using a buffer composed of 10 mM sodium bicarbonate and 500 mM sodium chloride dissolved in water (pH 7.8). After centrifugation, 5 μ L of the clear supernatant was diluted in 50 mM ammonium bicarbonate and 0.1% (*w/v*) RapiGest surfactant (Waters Corporation, Milford, MA, USA) and incubated at 80 °C for 15 min. Proteins were reduced with dithiothreitol (DTT), alkylated with iodoacetamide (IAA) and digested overnight by the addition of trypsin at 37 °C. Samples were acidified with trifluoroacetic acid and desalted using Sep-Pak tC₁₈ cartridges (Waters Corporation). Purified peptides were lyophilized and reconstituted in water with 0.1% (*v/v*) formic acid (FA). Consecutive LC-MS measurements were performed using a nanoACQUITY ultra-performance LC system (Waters Corporation) coupled to a SYNAPT G2-S mass spectrometer (Waters Corporation). Peptides were loaded onto a reversed-phase column (HSS T3 300 μ m \times 100 mm, 1.8 μ m, Waters Corporation, column temperature 55 °C) and separated at a flow rate of 8 μ L/min over 15 min using a gradient of 1–36% solvent B, which was acetonitrile with 0.1% (*v/v*) FA, while solvent A was water with 0.1% (*v/v*) FA. Dimethyl sulfoxide was added to the mobile phase after the column, as previously described [45]. Mass spectra were acquired in data-independent mode by MS^E [46]. Raw data were processed using ProteinLynx Global Server v.2.0.3 (Waters Corporation) and searched against a *T. aestivum* protein sequence database (UniProtKB release 2019_02, taxon ID: 4565, 142,700 entries). Postprocessing and LFQ was performed using ISOQuant v1.8 [47], applying a false-discovery rate cut-off of 0.01 at the peptide and protein level and Top3-based [48] protein quantification.

4.3. Statistical Analysis of Phenotypic Data

Details of the phenotypic data analysis were described in our previous study by Afzal et al. [18]. A total of 756 proteins were detected across the 148 wheat cultivars grown in three environments. However, only 114 proteins had a stable expression across all environments in at least one cultivar and met the following quality criteria: a heritability > 0.6, missing data < 20%, and expression in >50% and >80% of the cultivars in three and two environments, respectively.

Best Linear Unbiased Estimates (BLUEs) used in our GWAS analyses were estimated across the three environments (locations) assuming fixed genotypic effect in a mixed linear model. All phenotypic analyses were conducted using the statistical software R 3.3.2 [49] and software package ASReml-R 3.0 [50].

4.4. Genomic Data and Genotyping

Details of the genotyping approach were originally described by Rapp et al. [43]. Briefly, the cultivars were genotyped using the Diversity Arrays Technology (DArT) genotyping-by-sequencing platform (DArTseq). Two types of markers were delivered by DArTseq, codominant SNP (S) and dominant DArT (D) markers. SNP and DArT markers with more than 25% missing data and a minor allele frequency below 5% were omitted. The remaining missing values were imputed using the package LinkImpute [51]. A selection of 12,203 high-quality markers with a known genetic map positions and common chromosomal position between the wheat reference genome IWGSC RefSeq v1.0 and v2.1 were obtained. These polymorphic DArTseq markers were used for the subsequent genetic analyses. The physical map positions of the markers were retrieved from the Chinese Spring wheat genome (IWGSC CS RefSeq v2.1) implementing BLAST search by using NCBI database keeping the default parameter settings (https://www.ncbi.nlm.nih.gov/assembly/GCF_018294505.1, accessed on 15 February 2022). The default stringent Megablast algorithm did not produce any hits for some markers. We used BlastN algorithm for such markers, which allows a word-size down to seven bases. Implementing BLAST search using any of the two algorithms, when multiple hits for a given marker sequence were returned, only the top hit was kept because of its lowest e-value. In case of ties for the first position between two markers, the top hit was retrieved. In addition to using BLAST to determine positions of the DArTseq sequences, the sequences were mapped against the v2.1 IWGSC assembly [52]. DArTseq reads were first converted to fastq

format using custom Python scripts (v.3.9.12) [53]. Reads were then mapped to v2.1 IWGSC assembly [52] using bwa mem v 0.7.17 [54] with the default parameters and the -M flag. Samtools view v1.14 [55] was used to obtain mapped reads with quality > 30 (-q30) and primary alignments (-F256). Bam files were then sorted with samtools sort v1.14 [55] and transformed to bed format using bedtools bamtobed v2.29.2 [56]. Finally, mapped regions were compared to BLAST regions using bedtools intersect v2.29.2 [56].

4.5. Genome-Wide Association Study (GWAS)

Genome-wide association analysis was performed with the R package GAPIT (<http://zzzla.net/GAPIT>, accessed on 19 June 2022) [57] using a mixed linear model (MLM) accounting for population structure (Q) through a principal component (PC) analysis and for relationships between individuals through a kinship (K) matrix. Stringent selection of marker-trait associations (MTAs) was performed using the Bonferroni correction (with *p*-value cut-off at 0.05) to avoid false positives. The R package 'qqman' [58] was used to create the Manhattan plot representing all the MTAs. The genotypic variance explained by each quantitative trait locus (QTL) was calculated using a linear model fitting the significant MTAs ordered based on the strength of their association. The explained genotypic variance (pG) was calculated as: $pG = \frac{R^2_{adj}}{H^2}$ [59,60], where, R^2_{adj} is the adjusted R^2 from the linear model, and H^2 is the heritability of the trait. Only the most significantly associated markers were declared as putative QTLs and reported in the manuscript.

Box-and-whisker plots and barplots were generated using the R package 'ggplot2' [61] to examine the effect and the allele frequencies of QTL, respectively. We tested for significant differences between the groups of alleles using the *t*-test with 'stats' R package [49]. For the candidate gene search, the latest wheat reference genome (IWGSC RefSeq v2.1) and gene functional annotation information were downloaded from the URGI database [52]. High confidence (HC) genes located within the identified chromosomal regions of the significant QTLs were extracted. The genes with the functional annotation similar to the different domains of gluten and allergens were selected as potential candidate genes.

4.6. Verification of Two Major QTLs

Electrophoretic analysis was performed to identify HMW-GS composition of the experimental wheat cultivars. Crushed seeds of each cultivar were used to run the electrophoresis. HMW-GS were separated by sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which was carried out according to the protein extraction process of Singh et al. [62]. The HMW-GS were identified using the previously proposed nomenclature by Payne and Lawrence [63].

For marker-assisted selection for disease, a marker linked to the rust resistance gene Sr38-Yr1 was used. From the Illumina 25 K Infinium SNP array (SGS Institut Fresenius, TraitGenetics Section, Gatersleben, Germany), a codominant marker IAAV8501 linked with Sr38-Yr17-Lr37 locus (Paul Gruner pers. comm.) located on chromosome 2A (12,327,389 bp) was tested on more than 50% of the cultivars.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24021485/s1>.

Author Contributions: Conceptualization, C.F.H.L., S.T., D.S., K.E.H. and P.T.; methodology, C.F.H.L., S.T., D.S., K.E.H., M.A. and P.T.; data curation, M.S., V.C. and M.N.; formal analysis K.E.H., M.A., K.A.S. and P.T.; writing—original draft preparation, K.E.H., C.F.H.L. and P.T.; writing—review and editing, K.E.H., C.F.H.L., P.T., M.A., K.A.S., M.S., S.T. and D.S. All authors have read and agreed to the published version of the manuscript.

Funding: The project was supported by Deutsche Forschungsgemeinschaft (DFG) to C.F.H.L. (LO 1816-4/1), D.S. (Schu 646/17-1), and S.T. (TE599-3/1). K.E.H. was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program (BETTERWHEAT-FKZ: 2818405A18, 2818405B18).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The mass spectrometry proteomics data are available via ProteomeX-change with identifier PXD023654.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Shewry, P.R.; Hey, S. The contribution of wheat to human diet and health. *Food Energy Secur.* **2015**, *4*, 178–202. [[CrossRef](#)] [[PubMed](#)]
2. Goel, S.; Yadav, M.; Singh, K.; Jaat, R.S.; Singh, N.K. Exploring diverse wheat germplasm for novel alleles in HMW-GS for bread quality improvement. *J. Food Sci. Technol.* **2018**, *55*, 3257–3262. [[CrossRef](#)] [[PubMed](#)]
3. Shewry, P.R.; Halford, N.G.; Belton, P.S.; Tatham, A.S. The structure and properties of gluten: An elastic protein from wheat grain. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **2002**, *357*, 133–142. [[CrossRef](#)] [[PubMed](#)]
4. Payne, P.I.; Holt, L.M.; Worland, A.J.; Law, C.N. Structural and genetical studies on the high-molecular-weight subunits of wheat. Part 3. Telocentric mapping of the subunit genes on the long arms of homoeologous group 1 chromosomes. *Theor. Appl. Genet.* **1982**, *63*, 129–138. [[CrossRef](#)] [[PubMed](#)]
5. Kasarda, D.D. Glutenin polymers: The in vitro to in vivo transition. *Cereal Foods World* **1999**, *44*, 566–571.
6. Payne, P.I.; CorWeld, K.G.; Holt, L.M.; Blackman, J.A. Correlations between the inheritance of certain high molecular weight subunits of glutenin and bread-making quality in progenies of 6 crosses of bread wheat. *J. Sci. Food Agric.* **1981**, *32*, 51–60. [[CrossRef](#)]
7. Popineau, Y.; Cornec, M.; Lefebvre, J.; Marchylo, B. Influence of high M(r) glutenin subunits on glutenin polymers and rheological properties of gluteins and gluten subfractions of near-isogenic lines of wheat Sicco. *J. Cereal Sci.* **1994**, *19*, 231–241. [[CrossRef](#)]
8. Pirozi, M.; Margiotta, B.; Latiandra, D.; MacRitchie, F. Composition of polymeric proteins and bread-making quality of wheat lines with allelic HMW-GS differing in number of cysteines. *J. Cereal Sci.* **2008**, *48*, 117–122. [[CrossRef](#)]
9. Payne, P.I.; Corfield, K.G.; Blackman, J.A. Identification of a high-molecular-weight subunit of glutenin whose presence correlates with bread-making quality in wheats of related pedigree. *Theor. Appl. Genet.* **1979**, *55*, 153–159. [[CrossRef](#)]
10. Sapone, A.; Bai, J.C.; Ciacci, C.; Dolinsek, J.; Green, P.H.R.; Hadjivassiliou, M.; Kaukinen, K.; Rostami, K.; Sanders, D.S.; Schumann, M.; et al. Spectrum of gluten-related disorders: Consensus on new nomenclature and classification. *BMC Med.* **2012**, *10*, 13. [[CrossRef](#)]
11. Schuppan, D.; Pickert, G.; Ashfaq-Khan, M.; Zevallos, V. Non-celiac wheat sensitivity: Differential diagnosis, triggers and implications. *Best Pract. Res. Clin. Gastroenterol.* **2015**, *29*, 469–476. [[CrossRef](#)] [[PubMed](#)]
12. Zevallos, V.F.; Raker, V.; Tenzer, S.; Jimenez-Calvente, C.; Ashfaq-Khan, M.; Russel, N.; Pickert, G.; Schild, H.; Steinbrink, K.; Schuppan, D. Nutritional wheat amylase-trypsin inhibitors promote intestinal inflammation via activation of myeloid cells. *Gastroenterology* **2017**, *152*, 1100–1113.e1112. [[CrossRef](#)] [[PubMed](#)]
13. 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](#)] [[PubMed](#)]
14. EL Hassouni, K.; Sielaff, M.; Leiser, W.; Würschum, T.; Schuppan, D.; Tenzer, S.; Longin, C.F.H. Genetic Architecture Underlying the Expression of Eight Alpha-Amylase Trypsin Inhibitors. *Theor. Appl. Genet.* **2021**, *134*, 3427–3441. [[CrossRef](#)] [[PubMed](#)]
15. Jackson, E.A.; Holt, L.M.; Payne, P.I. Characterisation of high molecular weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. *Theor. Appl. Genet.* **1983**, *66*, 29–37. [[CrossRef](#)] [[PubMed](#)]
16. Canovas, F.M.; Dumas-Gaudot, E.; Recorbet, G.; Jorriin, J.; Mock, H.P.; Rossignol, M. Plant proteome analysis. *Proteomics* **2004**, *4*, 285–298. [[CrossRef](#)]
17. Afzal, M.; Pfannstiel, J.; Zimmermann, J.; Bischoff, S.C.; Würschum, T.; Longin, C.F.H. High-resolution proteomics reveals differences in the proteome of spelt and bread wheat flour representing targets for research on wheat sensitivities. *Sci. Rep.* **2020**, *10*, 1–11. [[CrossRef](#)]
18. Afzal, M.; Sielaff, M.; Curella, V.; Neerukonda, M.; El Hassouni, K.; Schuppan, D.; Tenzer, S.; Longin, C.F.H. Characterization of 150 wheat cultivars by LC-MS-based label-free quantitative proteomics unravels possibilities to design wheat better for baking quality and human health. *Plants* **2021**, *10*, 424. [[CrossRef](#)]
19. Mackay, T.F.C. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **2001**, *35*, 303–339. [[CrossRef](#)]
20. Holland, J.B. Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* **2007**, *10*, 156–161. [[CrossRef](#)]
21. Cooper, M.; van Eeuwijk, F.A.; Hammer, G.L.; Podlich, D.W.; Messina, C. Modeling QTL for complex traits: Detection and context for plant breeding. *Curr. Opin. Plant Biol.* **2009**, *12*, 231–240. [[CrossRef](#)] [[PubMed](#)]
22. Naz, A.A.; Kunert, A.; Lind, V.; Pillen, K.; Léon, J. AB-QTL analysis in winter wheat: II. Genetic analysis of seedling and field resistance against leaf rust in a wheat advanced backcross population. *Theor. Appl. Genet.* **2008**, *116*, 1095–1104. [[CrossRef](#)] [[PubMed](#)]

23. Lillemo, M.; Asalf, B.; Singh, R.P.; Huerta-Espino, J.; Chen, X.M.; He, Z.H.; Bjørnstad, A. The adult plant rust resistance loci Lr34/Yr18 and Lr46/Yr29 are important determinants of partial resistance to powdery mildew in bread wheat line Saar. *Theor. Appl. Genet.* **2008**, *116*, 1155–1166. [[CrossRef](#)] [[PubMed](#)]
24. Naz, A.A.; Dadshani, S.; Ballvora, A.; Pillen, K.; Léon, J. Genetic Analysis and Transfer of Favorable Exotic QTL Alleles for Grain Yield Across D Genome Using Two Advanced Backcross Wheat Populations. *Front. Plant Sci.* **2019**, *10*, 711. [[CrossRef](#)] [[PubMed](#)]
25. Chao, S.; Dubcovsky, J.; Dvorak, J.; Luo, M.-C.; Baenziger, S.P.; Matnyazov, R.; Clark, D.R.; E Talbert, L.; A Anderson, J.; Dreisigacker, S.; et al. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genom.* **2010**, *11*, 727. [[CrossRef](#)]
26. Jones, H.; Gosman, N.; Horsnell, R.; Rose, G.A.; Everest, L.A.; Bentley, A.R. Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: The *Aegilops tauschii* Coss. example. *Theor. Appl. Genet.* **2013**, *126*, 1793–1808. [[CrossRef](#)]
27. Ogonnaya, F.C.; Abdalla, O.; Mujeeb-Kazi, A.; Kazi, A.G.; Xu, S.S.; Gosman, N.; Lagudah, E.S.; Bonnett, D.; Sorrells, M.E.; Tsujimoto, H. Synthetic hexaploids: Harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.* **2013**, *37*, 35–122.
28. Rasheed, A.; Wen, W.; Gao, F.; Zhai, S.; Jin, H.; Liu, J.; Guo, Q.; Zhang, Y.; Dreisigacker, S.; Xia, X.; et al. Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* **2016**, *129*, 1843–1860. [[CrossRef](#)]
29. Rasheed, A.; Jin, H.; Xiao, Y.; Zhang, Y.; Hao, Y.; Zhang, Y.; Hickey, L.; Morgounov, A.; Xia, X.; He, Z. Allelic effects and variations for key bread-making quality genes in bread wheat using high-throughput molecular markers. *J. Cereal Sci.* **2019**, *85*, 305–309. [[CrossRef](#)]
30. Liu, S.; Chao, S.; Anderson, J.A. New DNA markers for high molecular weight glutenin subunits in wheat. *Theor. Appl. Genet.* **2008**, *118*, 177–183. [[CrossRef](#)]
31. Schwarz, G.; Sift, A.; Wenzel, G.; Mohler, V. DHPLC scoring of a SNP between promoter sequences of HMW glutenin x-type alleles at the Glu-D1 locus in wheat. *J. Agric. Food Chem.* **2003**, *51*, 4263–4267. [[CrossRef](#)] [[PubMed](#)]
32. Dreisigacker, S.; Xiao, Y.; Sehgal, D.; Guzman, C.; He, Z.; Xia, X.; Peña, R.J. SNP markers for low molecular glutenin subunits (LMW-GSs) at the *Glu-A3* and *Glu-B3* loci in bread wheat. *PLoS ONE* **2020**, *15*, e0233056. [[CrossRef](#)] [[PubMed](#)]
33. Bonafede, M.D.; Tranquilli, G.; Pflüger, L.A.; Peña, R.J.; Dubcovsky, J. Effect of allelic variation at the Glu-3/Gli-1 loci on breadmaking quality parameters in hexaploid wheat (*Triticum aestivum* L.). *J. Cereal Sci.* **2015**, *62*, 143–150. [[CrossRef](#)] [[PubMed](#)]
34. Afzal, M.; Sielaff, M.; Distler, U.; Schuppan, D.; Tenzer, S.; Longin, C.F.H. Reference proteomes of five wheat species as starting point for future design of cultivars with lower allergenic potential. *NPJ Sci. Food* **2022**. *submitted*. [[CrossRef](#)]
35. Zimmermann, J.; Hubel, P.; Pfannstiel, J.; Afzal, M.; Longin, C.F.H.; Hitzmann, B. Comprehensive proteome analysis of bread deciphering the allergenic potential of bread wheat, spelt and rye. *J. Proteomics.* **2021**, *247*, 104318. [[CrossRef](#)]
36. Juhász, A.; Belova, T.; Florides, C.G.; Maulis, C.; Fischer, I.; Gell, G.; Birinyi, Z.; Ong, J.; Keeble-Gagnère, G.; Maharajan, A.; et al. Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci. Adv.* **2018**, *4*, eaar8602. [[CrossRef](#)]
37. Mari, A.; Rasi, C.; Palazzo, P.; Scala, E. Allergen databases: Current status and perspectives. *Curr. Allergy Asthma Rep.* **2009**, *9*, 376–383. [[CrossRef](#)]
38. Bariana, H.S.; McIntosh, R.A. Characterisation and origin of rust and powdery mildew resistance genes in VPM1 wheat. *Euphytica* **1994**, *76*, 53–61. [[CrossRef](#)]
39. Seah, S.; Bariana, H.; Jahier, J.; Sivasithamparan, K.; Lagudah, E.S. The introgressed segment carrying rust resistance genes Yr17, Lr37 and Sr38 in wheat can be assayed by a cloned disease resistance gene-like sequence. *Theor. Appl. Genet.* **2001**, *102*, 600–605. [[CrossRef](#)]
40. Helguera, M.; Khan, I.A.; Kolmer, J.; Lijavetzky, D.; Zhong-qi, L.; Dubcovsky, J. PCR Assays for the Lr37-Yr17-Sr38 Cluster of Rust Resistance Genes and Their Use to Develop Isogenic Hard Red Spring Wheat Lines. *Crop. Sci.* **2003**, *43*, 1839–1847. [[CrossRef](#)]
41. Mondaini, A.; Rosyara, U.; Sehgal, D.; Dreisigacker, S. Selection signatures in the CIMMYT International Elite Spring and Semi-arid Wheat Yield Trials. *Plant Genome* **2021**, *15*, 1. [[CrossRef](#)] [[PubMed](#)]
42. Cruz, C.D.; Peterson, G.L.; Bockus, W.W.; Kankanala, P.; Dubcovsky, J.; Jordan, K.W.; Akhunov, E.; Chumley, F.; Baldelomar, F.D.; Valent, B. The 2NS translocation from *Aegilops ventricosa* confers resistance to the *Triticum* pathotype of *Magnaporthe oryzae*. *Crop. Sci.* **2016**, *56*, 990–1000. [[CrossRef](#)] [[PubMed](#)]
43. Rapp, M.; Schwadorf, K.; Leiser, W.L.; Würschum, T.; Longin, C.F.H. Assessing the variation and genetic architecture of asparagine content in wheat: What can plant breeding contribute to a reduction in the acrylamide precursor? *Theor. Appl. Genet.* **2018**, *131*, 2427–2437. [[CrossRef](#)]
44. Deutsch, E.W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J.J.; Kundu, D.J.; García-Seisdedos, D.; Jarnuczak, A.F.; Hewapathirana, S.; Pullman, B.S.; et al. The ProteomeXchange Consortium in 2020: Enabling ‘Big Data’ Approaches in Proteomics. *Nucleic Acids Res.* **2020**, *48*, D1145–D1152. [[CrossRef](#)] [[PubMed](#)]
45. Distler, U.; Łacki, M.K.; Schumann, S.; Wanninger, M.; Tenzer, S. Enhancing Sensitivity of Microflow-Based Bottom-Up Proteomics through Postcolumn Solvent Addition. *Anal. Chem.* **2019**, *91*, 7510–7515. [[CrossRef](#)] [[PubMed](#)]
46. Bateman, R.H.; Carruthers, R.; Hoyes, J.B.; Jones, C.; Langridge, J.I.; Millar, A.; Vissers, J.P.C. A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 792–803. [[CrossRef](#)]

47. Distler, U.; Kuharev, J.; Navarro, P.; Levin, Y.; Schild, H.; Tenzer, S. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **2013**, *11*, 167–170. [[CrossRef](#)]
48. Silva, J.C.; Gorenstein, M.V.; Li, G.-Z.; Vissers, J.P.C.; Geromanos, S.J. Absolute Quantification of Proteins by LCMSE. *Mol. Cell. Proteomics* **2006**, *5*, 144–156. [[CrossRef](#)]
49. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
50. Gilmour, A.R.; Gogel, B.J.; Cullis, B.R.; Thompson, R. *ASReml User Guide Release 3.0*; VSN International Ltd.: Hemel Hempstead, UK, 2009.
51. Money, D.; Gardner, K.; Migicovsky, Z.; Schwaninger, H.; Zhong, G.Y.; Myles, S. LinkImpute: Fast and accurate genotype imputation for non-model organisms. *G3 Genes Genom. Genet.* **2015**, *5*, 2383–2390.
52. Zhu, T.; Wang, L.; Rimbart, H.; Rodriguez, J.C.; Deal, K.R.; De Oliveira, R.; Choulet, F.; Keeble-Gagnère, G.; Tibbits, J.; Rogers, J.; et al. Optical maps refine the bread wheat *Triticum aestivum* cv Chinese Spring genome assembly. *Plant J.* **2021**, *107*, 303–314. [[CrossRef](#)]
53. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
54. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)] [[PubMed](#)]
55. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [[CrossRef](#)] [[PubMed](#)]
56. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)]
57. Wang, J.; Zhang, Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genom. Proteom. Bioinform.* **2021**, *19*, 629–640. [[CrossRef](#)] [[PubMed](#)]
58. Turner, S. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **2018**, *3*, 731. [[CrossRef](#)]
59. Utz, H.F.; Melchinger, A.E.; Schön, C.C. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* **2000**, *154*, 1839–1849. [[CrossRef](#)]
60. Würschum, T.; Leiser, W.L.; Kazman, E.; Longin, C.F.H. Genetic control of protein content and sedimentation volume in European winter wheat cultivars. *Theor. Appl. Genet.* **2016**, *129*, 1685–1696. [[CrossRef](#)]
61. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009; pp. 115–137.
62. Singh, N.K.; Shepherd, K.W.; Cornish, G.B. A simplified SDS-PAGE procedure for separating LMW subunits of glutenin. *J. Cereal Sci.* **1991**, *14*, 203–208. [[CrossRef](#)]
63. Payne, P.I.; Lawrence, G.J. Catalogue of alleles for the complex gene loci, Glu-A1, Glu-B1, and Glu-D1 which code for high-molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Res. Commun.* **1983**, *11*, 29–35.

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.