



Article

# Genome-Wide Association Analysis for Hybrid Breeding in Wheat

Monika Mokrzycka <sup>1</sup>, Stefan Stojalowski <sup>2</sup>, Mirosław Tyrka <sup>3</sup>, Przemysław Matysik <sup>4</sup>, Barbara Żmijewska <sup>4</sup>, Rafał Marcinkowski <sup>4</sup>, Urszula Woźna-Pawlak <sup>5</sup>, Róża Martofel <sup>5</sup>, Michał Rokicki <sup>5,6</sup>, Monika Rakoczy-Trojanowska <sup>7</sup> and Paweł Krajewski <sup>1,\*</sup>

<sup>1</sup> Institute of Plant Genetics, Polish Academy of Science, Strzeszyńska 34, 60-479 Poznań, Poland

<sup>2</sup> Department of Plant Genetics, Breeding and Biotechnology, West Pomeranian University of Technology Szczecin, Słowackiego 17, 71-434 Szczecin, Poland

<sup>3</sup> Department of Biotechnology and Bioinformatics, Faculty of Chemistry, Rzeszów University of Technology, Powstańców Warszawy 6, 35-959 Rzeszów, Poland

<sup>4</sup> Plant Breeding Strzelce Group IHAR Ltd., Główna 20, 99-307 Strzelce, Poland

<sup>5</sup> Poznań Plant Breeding Ltd., Kasztanowa 5, 63-004 Tulce, Poland

<sup>6</sup> Plant Breeding and Acclimatization Institute, Radzików, 05-870 Błonie, Poland

<sup>7</sup> Department of Plant Genetics, Breeding and Biotechnology, Institute of Biology, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warszawa, Poland

\* Correspondence: pkra@igr.poznan.pl

**Abstract:** Disclosure of markers that are significantly associated with plant traits can help develop new varieties with desirable properties. This study determined the genome-wide associations based on DArTseq markers for six agronomic traits assessed in eight environments for wheat. Moreover, the association study for heterosis and analysis of the effects of markers grouped by linkage disequilibrium were performed based on mean values over all experiments. All results were validated using data from post-registration trials. GWAS revealed 1273 single nucleotide polymorphisms with biologically significant effects. Most polymorphisms were predicted to be modifiers of protein translation, with only two having a more pronounced effect. Markers significantly associated with the considered set of features were clustered within chromosomes based on linkage disequilibrium in 327 LD blocks. A GWAS for heterosis revealed 1261 markers with significant effects.

**Keywords:** genome-wide association study; wheat hybrid breeding



**Citation:** Mokrzycka, M.; Stojalowski, S.; Tyrka, M.; Matysik, P.; Żmijewska, B.; Marcinkowski, R.; Woźna-Pawlak, U.; Martofel, R.; Rokicki, M.; Rakoczy-Trojanowska, M.; et al. Genome-Wide Association Analysis for Hybrid Breeding in Wheat. *Int. J. Mol. Sci.* **2022**, *23*, 15321. <https://doi.org/10.3390/ijms232315321>

Academic Editors: Andrés J. Cortés and Hai Du

Received: 19 October 2022

Accepted: 27 November 2022

Published: 5 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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* L.) is an important source of calories and proteins for humans worldwide [1]. This crop is cultivated on approximately 216 million hectares and yields over 765 million tons of grain [2]. For the last 60 years, the average yield of wheat has been increased 3-fold to the current level of approx. 3.5 t/ha [2,3], but the progress in this regard has been relatively slow in recent years. Norman Borlaug's "Green Revolution" initiated this continuous progress, which aimed to release new highly productive cultivars for developing countries. The use of heterosis and hybrid wheat breeding is a possible option for maintaining or boosting yield productivity in modern wheat cultivars. The possibility of the commercial use of wheat F1 hybrids was suggested as early as 1963 by Briggles [4]. Wheat hybrids may significantly outperform traditional cultivars regarding yield level (usually by under 5%, but in particular situations by more than 40%; [5]) because of the effect of heterosis and can reveal higher stability in various environments due to their heterozygotic nature [6–8]. Moreover, the certified hybrid seeds desired by producers protect the interests of breeders [9].

The availability of high-throughput genotyping techniques, such as genotyping by sequencing (GBS) or single nucleotide polymorphism (SNP) arrays, has accelerated global

genetic analyses of large genome species, such as wheat [10–14]. The results of genome-wide association studies (GWAS) can be applied to the selection of the most valuable genotypes (genomic selection). GWAS was recently used to analyze anther extrusion in wheat [15,16]. Association studies based on various types of populations for the wheat plant traits analyzed in this report were also performed; we refer to a number of them in Discussion, Sections 3.1–3.3. More complex analyses focusing on the multi-trait requirements of hybrid breeding have not yet been reported. In the breeding of traditional cultivars, the target of GWAS is relatively clear, and the objective of genomic selection is to identify genotypes with alleles associated with increased values of yield-related traits that are frequently affected by environmental changes [17–20]. The application of GWAS for genomic selection in hybrid breeding procedures is complex. The successful release of a cultivar results from selecting a proper maternal line (seed parent), paternal line (pollen parent), and prediction of the heterosis effect revealed by the final hybrid. The challenge is that a desirable trait in one of these three components can be undesirable in another. Furthermore, the best performance of the final hybrid should be achieved when numerous loci of high-yielding parents remain heterozygotic. Genomic selection in hybrid breeding should be adjusted to meet these requirements.

In addition to the agronomic performance of related traits, the condition of heterosis breeding success is the appropriate flowering biology of the parental components. Different elements of flowering biology determine the suitability of male and female wheat parents for hybrid breeding. Efficient anther extrusion and a high quantity of released pollen over an extended period are considered the most important traits desired in male candidates for the successful production of hybrid wheat seeds. For female parents as pollen recipients, more significant is the ability to open flowers in the time necessary to get pollinated for extended periods, coupled with hairy stigmas. Considering the plant height, it is recommended that the pollen parent be taller than the seed parent because the pollen of wheat is much less mobile than the pollen of typical wind-pollinated plants such as maize and rye [3,7,16,21–23].

Despite several studies, the processes of flower opening and anther extrusion, a very complex and dynamic phenomenon combining anatomical, physiological, and biochemical parameters, are still not fully understood or defined [24–26]. Nevertheless, the variation and heritability of flowering-related traits are considered moderate to high, which provides the opportunity to select desirable parental components [7,27].

In the presented work, we use the data presented in [28], which was applied to the analysis of the genetic structure of a large pool of 509 wheat varieties and breeding lines. On adding to them the phenotypic observations obtained in two independent series of field trials, we presume that by using GWAS we will identify chromosomal regions associated with six agronomically important traits that are significant for wheat hybrid breeding.

To verify this presumption, we set the following objectives: (1) to apply the genotyping data [28] consisting of SNP marker observations determined by the DArTseq platform for GWAS, and (2) to identify chromosomal regions associated with six traits within a set of wheat genotypes across different environments.

## 2. Results

### 2.1. Phenotypic Characteristics of the Accessions

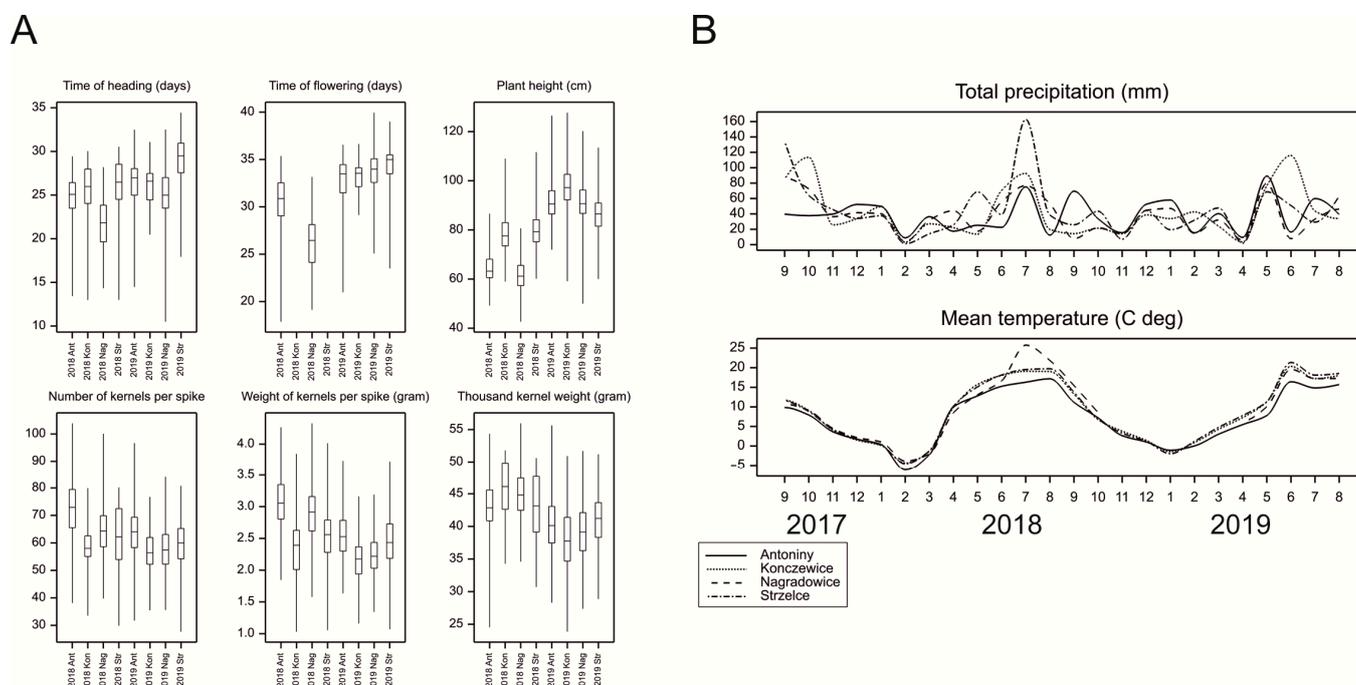
Six phenotypic traits were assessed in eight environments. The distributions of genotypic means (BLUPs obtained in individual experiments; Table S1) were approximately normal; a slight skewness to the right was observed for time of flowering—FT (Figure S1). Two phenological traits, time of heading—HT and FT, were positively correlated with each other and with plant height—PH ( $p$ -value < 0.05; Figure S1). In contrast, for PH, a negative correlation was observed with yield-related traits (number of kernels per spike—KN, weight of kernels per spike—KW, and thousand kernel weight—TKW). KW was positively correlated with TKW.

In ANOVA, all fixed effects, i.e., effects of the year (Y), localization (L), and Y × L interaction, were significant (at least at a *p*-value < 0.05, but most of them at *p*-value < 0.001). The variance components for random effects of genotypes (G) were much larger than those for G × Y and G × L interactions for traits HT, FT, and PH, which was reflected by the large heritability for these traits (88–90%; Table 1). The variance components were in the same order for yield-related traits, and therefore, heritability was smaller (38–54%).

**Table 1.** Mean values, coefficients of variation (CV), variance components (with standard errors) for genotypes (G) and their interactions with year (Y) and localization (L), and broad-sense heritability coefficients for the six traits.

Trait	Mean	Minimum	Maximum	CV	Variance Component ± s.e.				Heritability (%)
					G	G × Y	G × L	Error	
HT	25.45	10.5	34.44	13.8	6.06 ± 0.42	0.91 ± 0.07	0.36 ± 0.03	1.56 ± 0.03	90.41
FT	31.67	17.87	39.95	11.74	4.04 ± 0.32	1.39 ± 0.11	0.46 ± 0.04	1.38 ± 0.03	80.77
PH	81.5	42.66	127.73	17.42	32.52 ± 2.37	5.92 ± 0.55	2.41 ± 0.31	22.15 ± 0.42	86.79
KN	62.24	27.69	103.9	15.88	14.05 ± 1.73	9.68 ± 1.12	11.44 ± 1.03	63.5 ± 1.2	54.60
KW	2.53	1.03	4.33	19.49	0.018 ± 0.003	0.028 ± 0.003	0.023 ± 0.002	0.14 ± 0.003	38.79
TKW	41.87	23.86	55.96	12.03	4.36 ± 0.56	5.41 ± 0.42	2.25 ± 0.18	9.72 ± 0.18	52.97

A comparison of the distributions of genotypic means in particular experiments showed that accessions headed and flowered earlier and were shorter in 2018 than in 2019 (Figure 1A), and that the yield parameters were higher in 2018. The weather conditions, as characterized by monthly precipitation and temperatures, differed slightly between the two growing seasons. In 2017/2018, winter temperatures were lower, and summer precipitation was slightly higher than in 2018/2019 (Figure 1B).



**Figure 1.** (A) Distributions of genotypic means in experiments; Ant—Antonyń, Kon—Konczewice, Nag—Nagradowice, Str—Strzelce; (B) Observations of weather parameters in 2017–2019 at four locations of experiments (ten-day summaries).

Stability analysis in the AMMI model revealed that, in general, the reactions of genotypes to conditions in the four locations were more similar in 2019 than in 2018 (Figure S2A). In 2018, for traits HT, PH, KN, and TKW, a considerable difference existed between reactions to conditions in (Antoniny, NAGRADOWICE) vs. (KONCZEWICE, STRZELCE). The patterns of the interactions for localization were similar for HT and TKW. Genotypes showed continuous distributions of interactions. For FT accessions, PHR\_6 (Akteur), STH\_148 (UKR 12), and for KW accessions, STH\_65 (STH 9025), STH\_167 (Cornea ost.) showed outlying instability. The largest significant correlations existed between instability variances for genotypes estimated for pairs of traits: HT and FT and KW and TKW (Figure S2B,C); correlations for trait pairs (PH, FT) and (KW, KN) were statistically significant but very small. The ranks of genotypes concerning instability variances for all traits are presented in Table S2; they show that the most stable genotypes for one trait were usually much less stable for the other.

## 2.2. Association Analysis

GWAS revealed 7603 SNPs with polymorphisms significantly associated with at least one phenotypic trait (only SNPs with more than five accessions in each of the two homozygous classes; BH corrected  $p$ -value < 0.05). The characteristics of the allelic substitution effects for these loci are shown in Table 2.

**Table 2.** Characteristics of allelic substitution effects for SNPs associated significantly with phenotypic traits.

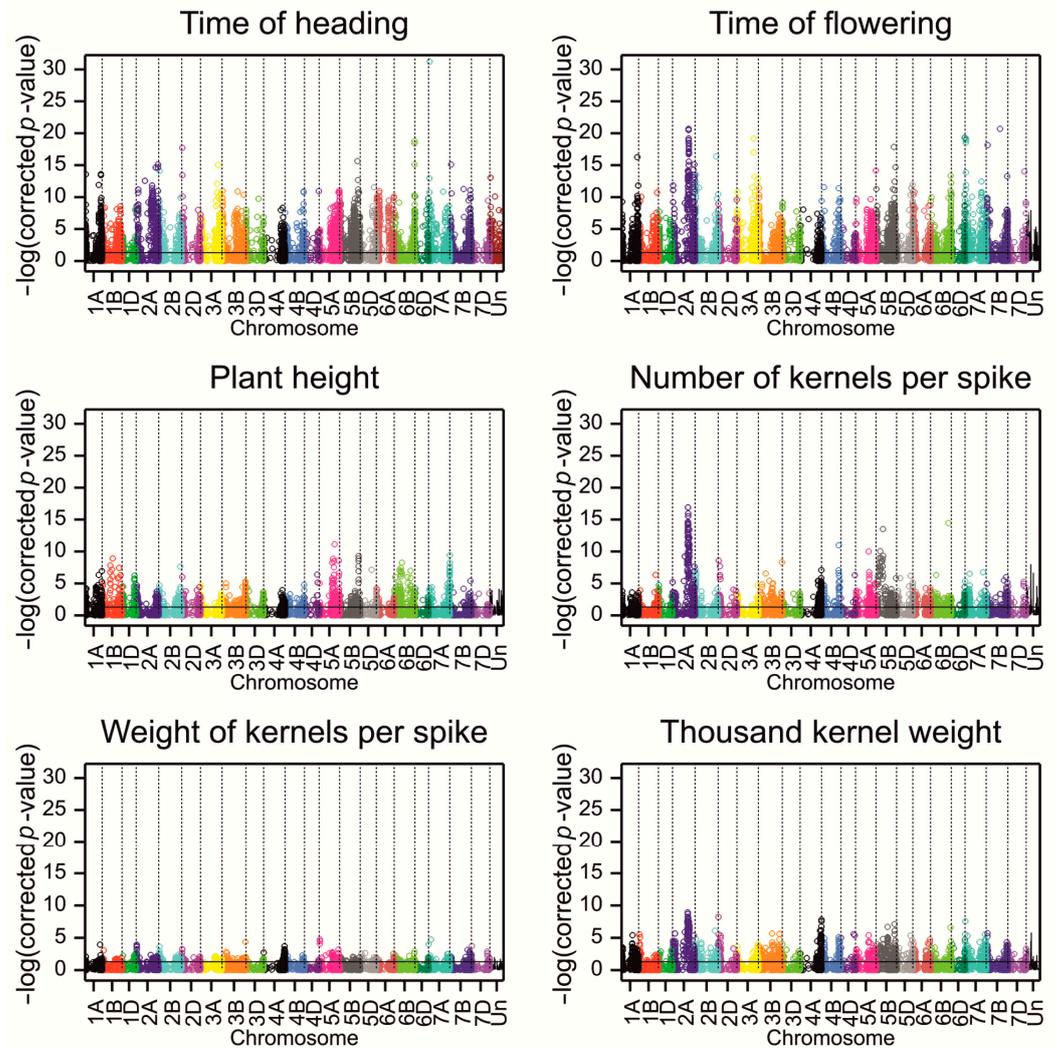
Trait	Number of Associations	Negative Significant Effects		Positive Significant Effects		2th Percentile	
		Min	Max	Min	Max	Lower	Upper
HT	4206	−1.57	−0.0024	0.0003	1.49	−0.68	0.72
FT	4108	−1.30	−0.0027	0.0014	1.13	−0.61	0.64
PH	1999	−4.24	−0.0006	0.0013	3.18	−1.72	1.63
KN	1858	−4.29	−0.0015	0.0113	3.27	−1.36	1.54
KW	738	−0.09	−0.0001	0.0002	0.12	−0.05	0.05
TKW	1908	−1.41	−0.0002	0.0001	1.19	−0.66	0.61

By considering also the criterion related to the size of the substitution effect being in the second lower or upper percentile, the number of associated markers was reduced to 1273, and the number of significant associations was reduced to 1344 (Figure 2, Table 3 and Table S3).

The mean linkage disequilibrium (LD) with other SNPs in the region  $\pm 5$  Mb was determined for significant SNPs. A detailed analysis of LD has been provided in [28] (a heatmap of LD on chromosome 1A in Figure S2A, exemplary LD clusters in Figure S2B, and the plots of LD vs. physical distance between markers with 0–20 Mb distance intervals in Figure 5A of [28]). The percentage of associations with a small mean LD (<0.01) in this interval varied from 5% for PH to 20% for KW (Table 3). The significant SNPs were clustered into small groups of no more than five markers within chromosomes based on LD (Table 3). The highest number of LD clusters was on chromosome 3A for the two phenological traits, on chromosome 2B for PH, on 2B and 6B for KN, on 7B for KW, and on 2A for TKW (Figure 3).

The fraction of associations showing interaction with the environment was the lowest for PH and KN and the highest for KW. Associations were unevenly distributed in the wheat subgenomes. Most of the associated SNPs mapped to the A and B subgenomes (38.4% and 45.01%, respectively), whereas only 14.14% were derived from D, and 2.44% remained unmapped (Table 3, Figure 2); however, the proportion of the relative number of SNPs was similar: 8.99% from the A subgenome, 10.14% from B, 8.45% from D and 11.07 from Un. For five traits, most associations were detected in subgenome A, and only for PH were most associations detected in subgenome B. More than 50% of the associated SNPs for all the traits were located in genes. Most SNPs associated with traits were predicted to modify protein translation, and only two SNPs significantly influenced

the protein structure. The first SNP (1130302|F|0-44|TG), which was associated with FT, altered the splicing donor site in TraesCS6A02G085100 (Arabidopsis RPT3 ortholog), which is responsible for ATP-dependent degradation of ubiquitinated proteins. The second SNP (1090593|F|0-44|CT) was associated with TKW and introduced a STOP codon in the gene TraesCS4B02G086500, annotated with “carbohydrate metabolic process” and “polygalacturonase activity” GO terms.



**Figure 2.** Manhattan plots for associations of SNPs with studied phenotypic traits. Points represent  $p$ -values for associations corrected by the Benjamini–Hochberg method. The black line visualizes the critical significance level of 0.05.

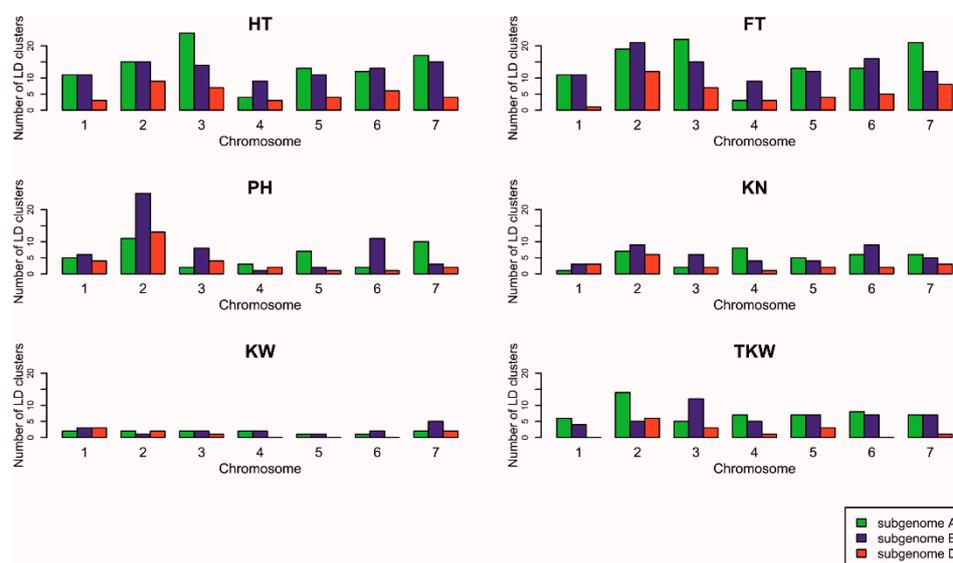
Allelic substitution effects for different traits were correlated (Figure S3). However, not all correlations were of the same size and sign as those of the traits themselves. For example, allelic effects for PH were negatively correlated with effects for phenological traits (positive correlation for traits), and effects for KW were positively correlated with phenological traits (negative correlation for traits).

SNPs associated with traits were divided into three sets corresponding to plant phenology (HT or FT, set 1), PH (set 2), and kernel properties (KN or KW or TKW, set 3), with the numbers of associations in these sets being 592, 391, and 429, respectively. There were 26, 99, and 26 SNPs common to sets 1 and 2, 1 and 3, and 2 and 3, respectively, and the 12 SNPs common for all three sets (Figure 4, Table S3). For pairs of traits, the largest number of common significant SNPs was recorded for HT and FT (312), which is consistent with a large correlation between these two traits reported above.

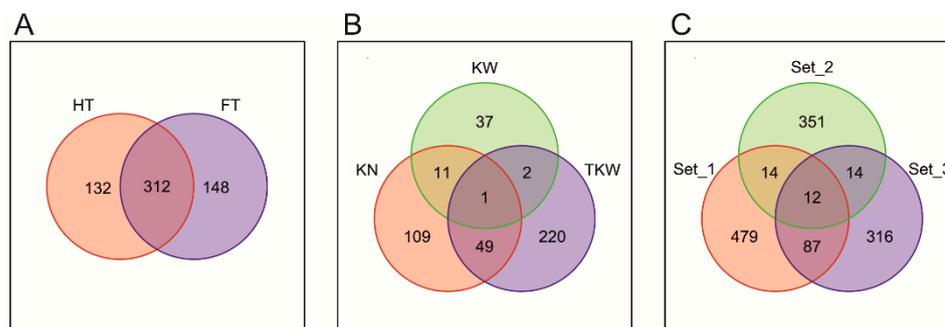
**Table 3.** GE interaction and distribution of trait-associated SNPs in linkage blocks, subgenomes, and genes.

Trait	Number of Associations	Number of SNPs with Mean LD <0.01	Number of SNP Clusters	With GE Interaction (% out of Significant)	% in Subgenome			% in Genes	Number of SNPs Affecting Protein Translation			
					A	B	D		High	Low	Moderate	Modifier *
HT	444	71	220	66.0	47.1	36.9	13.3	55.4	0	57	45	342 (144)
FT	460	78	238	66.1	43.7	34.8	18.7	56.7	1	64	38	357 (158)
PH	391	21	123	38.1	20.2	63.2	13.3	59.6	0	52	66	273 (115)
KN	170	25	94	31.2	44.1	32.4	21.8	55.9	0	18	20	132 (57)
KW	51	10	36	68.6	41.2	39.2	19.6	54.9	0	7	6	38 (15)
TKW	272	22	115	61.4	55.1	36.4	5.9	58.5	1	23	29	219 (106)

\* Total number of markers indicated by VEP [29] (in brackets: number of markers assigned to specific genes).



**Figure 3.** Number of LD clusters containing associated SNPs by chromosomes for each trait.



**Figure 4.** The number of common and specific SNPs in (A,B) association sets for traits and (C) association set 1 (HT or FT), set 2 (PH) and set 3 (KN, KW or TKW).

Of the 26 markers that belonged to sets 1 and 2, 17 were assigned to genes, with 5 being modifiers and 4 substitutions having moderate effects (Table 4). There were three markers with concordant effects on earliness and plant height, with the modifying effect of substitutions. Two of these (2253029|F|0-10|CT and 1237800|F|0-13|CG) were associated with genes (Table S3). Out of 12 SNPs common for sets 1, 2 and 3, eleven were associated

with KN and one with TKW. The same direction of effect could be seen for FT and KW, while the opposite direction was seen for KN and PH.

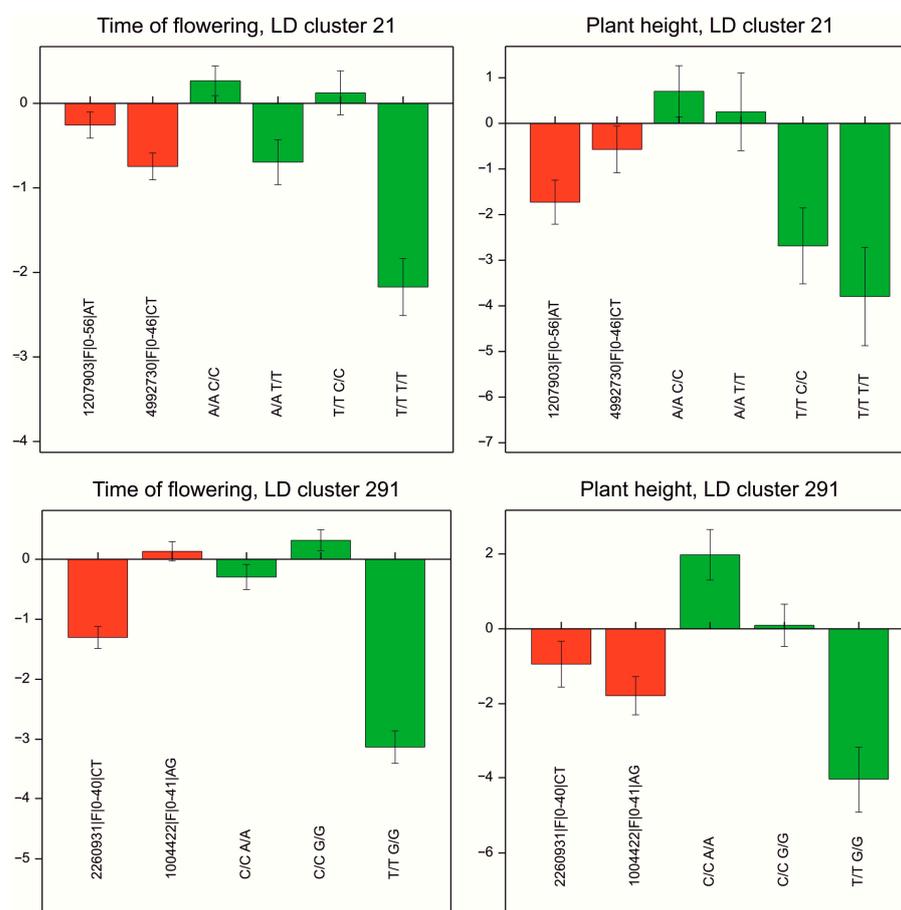
**Table 4.** Markers with concordant effects for sets 1 and 2 (in bold) and assigned to genes. Types of translation effects caused by substitution are provided (LOW, MDR—moderate, or MFI—modifier).

Marker	Gene	Effect	Interpro Description
<b>2253029</b>   F   0-10   CT (negative)	TraesCS2A02G482200	MFI	NAD-dependent epimerase/dehydratase; NAD(P)-binding domain superfamily
3938110   F   0-10   CG	TraesCS2B02G045100	LOW	NB-ARC;P-loop containing nucleoside triphosphate hydrolase; Leucine-rich repeat domain superfamily
<b>1237800</b>   F   0-13   CG (positive)	TraesCS2B02G164700	MFI	F-box-like domain superfamily
4910338   F   0-15   GC	TraesCS2B02G475700	MDR	Zf-FLZ domain; Zf-FLZ domain
2322929   F   0-10   AT	TraesCS2B02G490600	MFI	Ribonuclease H-like superfamily; Exonuclease, RNase T/DNA polymerase III; Ribonuclease H superfamily
2322355   F   0-40   CG	TraesCS2B02G521100	LOW	Glycosyl transferase;1,3-beta-glucan synthase subunit FKS1-like, domain-1
1238701   F   0-16   GA	TraesCS2D02G127300	LOW	F-box domain;F-box-like domain superfamily
1675478   F   0-15   TG	TraesCS3A02G506600	LOW	NAD(P)-binding domain superfamily; NAD(P)-binding domain superfamily
7350269   F   0-8   TC	TraesCS3A02G517700	LOW	
1049114   F   0-20   AG	TraesCS3D02G511900	MFI	Ubiquitin-like domain; Ubiquitin domain; Ubiquitin-like domain superfamily
1675534   F   0-35   AC	TraesCS3D02G513900	MDR	UDP-glucuronosyl/UDP- glucosyltransferase
2252787   F   0-19   CT	TraesCS3D02G529700	MFI	Coenzyme Q-binding protein COQ10, START domain; START-like domain superfamily
7352843   F   0-43   AT	TraesCS3D02G531000	MDR	Transcription initiation factor IIA, gamma subunit; Transcription factor IIA, helical; Transcription factor IIA, beta-barrel; Transcription initiation factor IIA, gamma subunit, C-terminal
7352096   F   0-35   GC	TraesCS3D02G541900	LOW	Uncharacterised conserved protein UCP015417, vWA
7353078   F   0-15   CA	TraesCS3D02G542800	MDR	Oxoglutarate/iron-dependent dioxygenase; Non-haem dioxygenase N-terminal domain; Isopenicillin N synthase-like
3024403   F   0-23   CT	TraesCS3D02G543100	LOW	DnaJ domain; Tetratricopeptide-like helical domain superfamily; Tetratricopeptide repeat-containing domain
1009915   F   0-65   CG	TraesCSU02G059500	LOW	Leucine-rich repeat, cysteine-containing subtype; SKP1/BTB/POZ domain superfamily; BTB/Kelch-associated; Leucine-rich repeat domain superfamily
1056528   F   0-9   TG (negative) *	-	MFI	-

\* marker not in the gene but showing congruent association with HT, FT, and PH.

### 2.3. Marker LD Clusters

Clusters of markers significantly associated with HT, FT (set 1), and with PH (set 2) (Table 3) were considered for LD-based GWAS. Combinations of their genotypes, represented by at least 25 accessions from the investigated collection, were considered for further analysis. The characteristics of the 327 marker clusters are presented in Table S4. Among them, we attempted to identify the effects of genotypic combinations that were higher than the separate allele effects of particular single markers. Examples of genotypic combinations with large negative effects on FT, HT, and PH are shown in Figure 5. For marker LD block no. 21, which clusters markers “1207903|F|0-56|AT\_4992730|F|0-46|CT”, the variant “T/T T/T” was present in 26 accessions and was related to early flowering (effect of  $-2.9$  days) and low plants (effect of  $-3.8$  cm) in comparison to the average of all accessions. One of these markers was assigned as an upstream modifier of the gene *TraesCS1B02G028100*, annotated as involved in “serine-type endopeptidase inhibitor activity.” Other interesting marker clusters with higher effects contributed by haplotypes were identified, for example, no. “291”, 2260931|F|0-40|CT\_1004422|F|0-41|AG. We noted that 76% of the genotypic variants in marker LD blocks were fully homozygous. Therefore, their information was equivalent to the knowledge of haplotypes.



**Figure 5.** Effects found in the association analysis: red—allelic effects of individual SNPs, averaged over eight experiments; green—haplotype effects (deviations from the general mean). Bars represent standard errors of effects. Illustrated are the results for two marker clusters with effects concordant for earliness (FT) and plant height (PH).

### 2.4. Heterosis

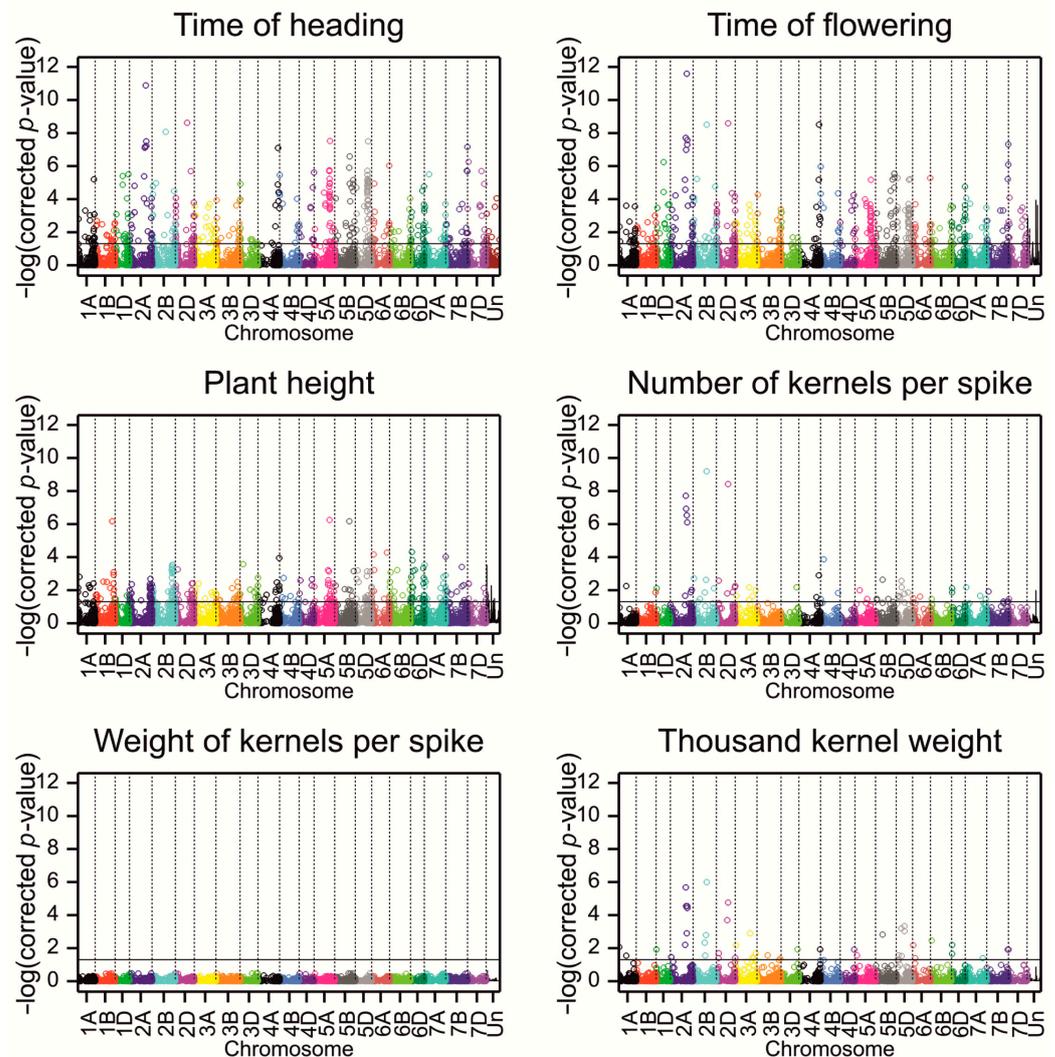
Searching for heterosis of SNP polymorphisms (on the set of SNPs with more than five heterozygous lines; BH corrected  $p$ -value  $< 0.05$ ) revealed 1261 SNPs with significant effects (Table 5 and Table S5), with the largest number of effects for phenological traits and

no effects for KW. Significant SNPs were evenly distributed over the subgenomes, with 53–80% of SNPs in the genes (Figure 6). None of the SNPs revealed a high translation effect, and the majority of the predicted effects were classified as low or modifying protein translation.

**Table 5.** Characteristics of heterosis effects for the five traits (KW was omitted due to the lack of significant effects).

Trait	Number of Significant Effects	% in Genome			% in Genes	Number SNPs with Protein Translation Effect (in Genes)		
		A	B	D		Low	Moderate	Modifier *
HT	437	33.9	28.6	35.5	57.7	106	52	275 (94)
FT	393	35.4	28.0	35.4	57.0	96	45	249 (83)
PH	324	35.5	31.8	28.7	52.8	52	41	229 (78)
KN	62	37.1	32.3	29.0	62.9	18	3	41 (18)
TKW	45	42.2	20.0	37.8	68.9	13	2	30 (16)

\* Total number of markers indicated by VEP [29] (in brackets: number of markers assigned to specific genes).



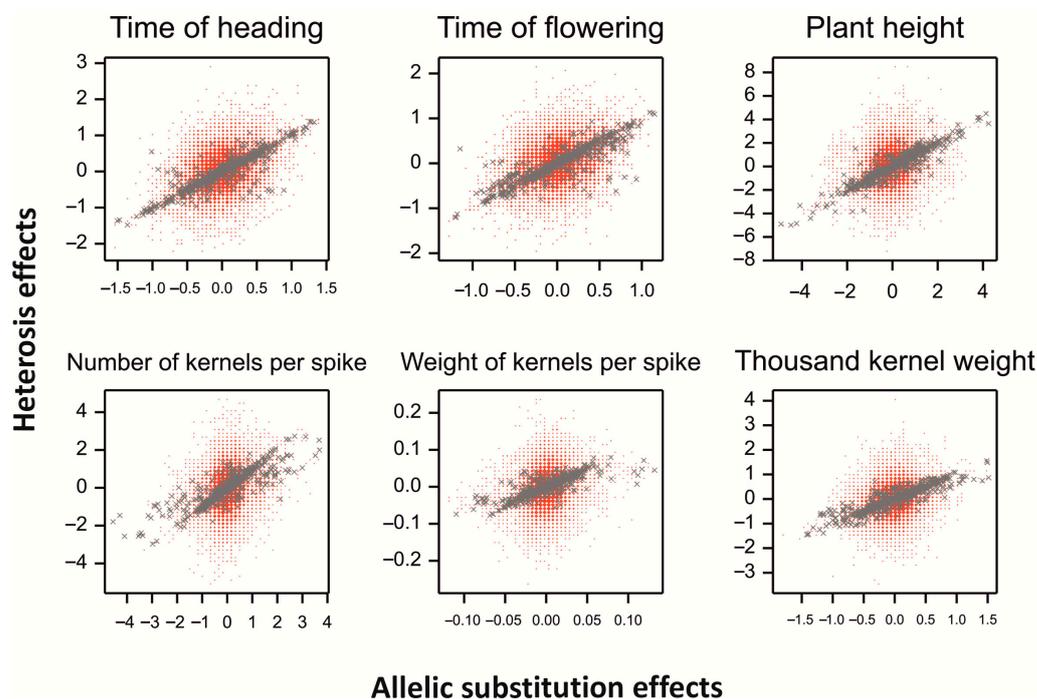
**Figure 6.** Manhattan plots for heterosis effects of SNPs. Points represent *p*-values for associations corrected by the Benjamini–Hochberg method. The black line visualizes the critical significance level of 0.05.

Association sets for traits (HT, FT), (PH), (KN, KW, and TKW) based on heterosis effects contained 480, 324, and 83 markers, respectively. The latter (yield-related) set contained 52 markers assigned to genes; the heterosis effects in this set were: from  $-4.50$  to  $3.09$  grains for KN, from  $-0.12$  to  $0.09$  g for KW, and from  $-2.09$  to  $4.13$  g for TKW. The highest negative effect for KN ( $-4.50$  grains) was for marker 3024735|F|0-10|GA (TraesCS7B02G481400).

The correlations of heterosis effects were similar to those of additive effects (Figure S4).

### 2.5. Allelic Substitution Effects vs. Heterosis Effects

Allelic effects were correlated with heterosis effects for markers with high heterozygosity (Figure 7). The range of additive effects was smaller than that of heterosis effects for all traits.



**Figure 7.** Heterosis effects vs. additive effects for six phenotypic traits; gray points for markers with heterozygosity  $> 0.5$ .

One marker, 1238701|F|0-16|GA (TraesCS2D02G127300), belonged to sets 1 and 2 with respect to additive effects (effects for HT  $-0.75$ , for PH  $2.47$ ), and to set 3 with respect to heterosis effects (effect  $2.25$  for TKW).

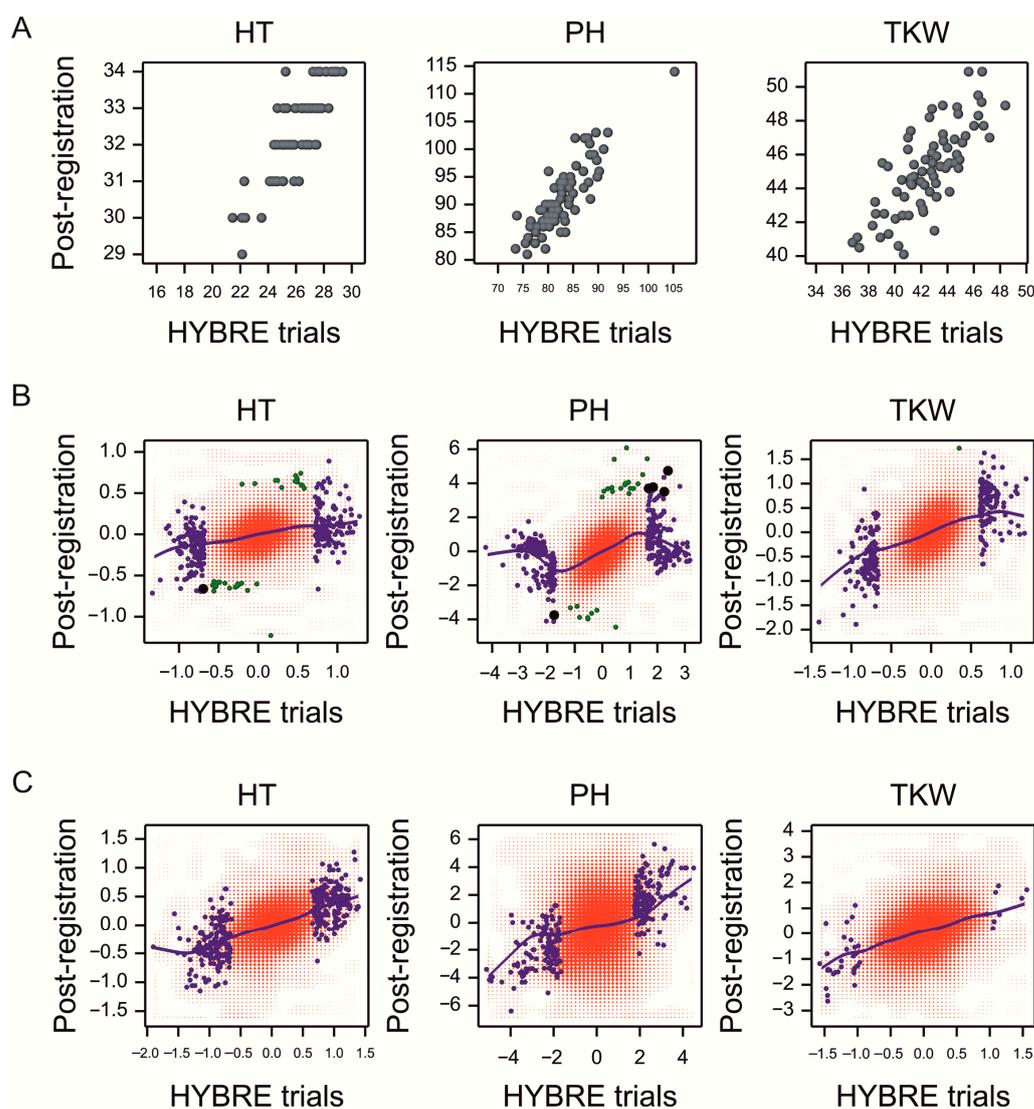
### 2.6. GO Annotation of Associated Markers

GO terms (biological processes) represented in sets of genes assigned to SNPs in association sets 1, 2, and 3 for allelic effects and heterosis effects are shown in Table S6A and Table S6B, respectively. Most frequently (16–26%), markers associated with both additive and heterosis effects were involved in redox processes and protein phosphorylation. The next most common GO terms were regulation of transcription (DNA-templated), carbohydrate metabolic process, and transmembrane transport. Responses to auxins were identified in groups of genes responsible for plant height additive effects more frequently than proteolysis.

### 2.7. Validation of Results Using Post-Registration Trial Data

Data from post-registration trials (PRT) were used to validate the GWAS results obtained in the HYBRE experiments. The general varietal means of HT, PH, and TKW from the

two series of experiments were correlated (0.84, 0.86, and 0.76, respectively;  $p$ -value < 0.001; Figure 8A). Allelic effects from GWAS were also correlated (0.29, 0.42, and 0.40, respectively;  $p$ -value < 0.001; Figure 8B). However, only three SNPs were significantly associated with traits in both series of trials: one for HT (3024420|F|0-9|CT, modifier, intergenic), and two for PH (1087592|F|0-37|GA, modifier, downstream of TraesCS5A02G277900 and 1126438|F|0-22|CT, modifier, intergenic). The heterosis effects were also correlated (Figure 8C).



**Figure 8.** (A) Scatterplots of varietal means observed in post-registration experiments (PRT) vs. means observed in HYBRE trials, for 75 varieties occurring in both series of experiments. (B) Density plots of allelic effects for markers estimated in post-registration experiments (PRT) vs. effects estimated in HYBRE trials in the set of 75 varieties. (C) Density plots of heterosis effects estimated in PRT vs. HYBRE. (B,C) Blue—significant in HYBRE, green—significant in PRT, and black—significant in both analyses. The blue lines are trend lines determined by smooth regression using the thin plate method.

### 3. Discussion

The main goal of hybrid wheat breeding is to exploit the heterosis effect. Final success depends on several factors, but yield is a crucial goal. Without a significant increase in productivity offered by hybrid cultivars, they will not be the choice of farmers; as an alternative, they may apply less expensive seeds of well-performing classic cultivars. The cost of hybrid seed production is also important. It depends on numerous factors related

to plant morphology and the biology of flowering [3,7,30]. The cross-pollination of wheat plants is based on the wind, but it is significantly less effective than that of typical open-pollinated plants, such as rye and maize. Waines and Hedge [31] noticed that singular wheat pollen could be found as far as 1000 m from the plantation, but hybrid seeds are usually undetectable beyond 20–30 m. The distance between parental lines of wheat allowing for economically sufficient effectiveness of seed production is limited to 2–3 m, and larger values can be applied when the pollen parent is taller than the seed parent. The reduction of plant height in wheat (highly recommended for maternal components) can be achieved by utilizing properly selected reduced height (Rht) genes.

In this study, we present a multifaceted statistical analysis aimed at providing information on a broad population of cultivars and breeding strains, which can primarily be used in wheat hybrid breeding. We investigated other morphological and phenological traits significant for efficient seed production and chosen traits related to wheat productivity, aiming to verify whether our non-standard genome-wide association study methods have potential application in hybrid breeding of wheat.

The analysis of variance showed that genotypic variability, measured in relation to variabilities caused by genotype by year and genotype by localization interactions, and, consequently, broad-sense heritability, was higher for phenological traits and for plant height than for yield-related traits. This indicates the possibility of selecting valuable parental forms independently, to some extent, from the target environment. Additionally, stability analysis (AMMI) showed that, although a given genotype may be subjected to various degrees of environmental variance for various traits, some correlation exists among instability variances for traits related to hybrid component selection.

However, our association analysis was performed assuming that SNP variant effects can also interact with environmental conditions. This aspect of GWAS analysis is often addressed by performing separate analyses for each environment (cf. [32–34]), mainly because of the limited options presented by the data analysis software. Based on the mixed model developed by Malosetti et al. [35], our approach allowed for the explicit testing of markers by environment interaction effects, which provided a conclusion on the lower environmental variability of SNP effects for plant height than for phenology. We also performed GWAS analysis using marker LD clusters. The most promising marker clusters were those for which the effects of genotypic combinations were higher than the allelic effects of individual markers. Furthermore, concordant effects of earliness and plant height are required for the practical importance of the haplotype. The applied procedure allowed the identification of groups of markers with a high impact on phenotypic traits, which could be expected for individual polymorphisms. Thus, we demonstrated the possibility of using the phenomenon of non-allelic interactions in LD-informed hybrid breeding. Moreover, as hybrid breeding is based on the effects of heterozygous materials, another association analysis was performed with respect to heterosis. It aimed to identify genomic loci for which heterozygotes could be more profitable than homozygotes, especially for yield-related traits. Although the results obtained in this way are far less valuable than those that could be gained from observations of hybrids, this is undoubtedly progress compared to the methodological approaches used to date in this type of research.

Validation of GWAS results can be done by an independent experimental study concerning another pool of genotypes. In our case, the pool covers most of the accessions that are practically interesting for project stakeholders. Another form of validation can be based on a real breeding process; this has been started by passing our results to breeders, who have already performed selected crosses and are currently assessing the value of hybrids by their standard procedures. In this report, to verify the results of GWAS, publicly available data from independent, post-registration trials performed in a wider set of environments were used. The mean BLUPs for genotypes from the HYBRE experiments were correlated with the general means from the PRT. Moreover, substantial correlations between allelic substitution effects obtained from GWAS on both datasets were obtained; the same was true for heterosis effects. The effects concerning plant height were found to be the most

correlated. This means that the data obtained in the HYBRE experiment are representative, and it can be assumed that new breeding creations can be evaluated in limited field experiments rather than in large experimental systems.

### 3.1. Flowering-Related QTLs and Genes

The expected female component of the hybrid should contain long stigmatic hairs that are fully extruded and receptive for long periods [7]. Wheat stigma remains receptive for up to 13 days after anthesis; however, it is usually highly receptive for no more than 3 days after anthesis [21]. Effective anther extrusion of the pollen parent is crucial for the successful setting of hybrid seeds. Recently, Denisow et al. [36] showed that the anther extrusion plays a much more important role in the contribution to the final amount of pollen available for cross-pollination than previously thought. Flowering generally begins at the center of a wheat spike and proceeds in both directions (up and down). Within each spikelet, the primary floret opens first, followed by the secondary, tertiary, etc. The first two florets of the spikelet produce the most abundant anthers, filled with the most viable pollen [37]. Thus, it is suggested that when these highly effective florets of the male component begin flowering, the female parent should be ready to receive pollen. The optimal difference between the flowering times of the seed parent and pollen parent is 2–3 days [30]. The phenotypic efficiency of some of the SNP markers indicated in our analyses met these criteria. The studied set of wheat accessions had been focused on genotypes cultivated in Europe [28]. Therefore, phenotypic variation was, in some way, limited. Despite this, some of the selected SNP markers revealed additive effects ranging from 1–1.6, corresponding to a difference from 2 to over 3 days of flowering time of homozygotic wheat genotypes, which are potential parental components of hybrids.

Markers associated with flowering time are necessary to synchronize the flowering of hybrid wheat components. Eight QTLs for heading time have been reported [38]. One of them, QHD\_7A\_psr\_ParW670\_CFLN17, corresponding to a relatively short region 710–719 Mbp on 7A, colocalized with the 3025631|F|0-10|GA marker found in our study. The regulation of flowering has been extensively studied at the gene level. Flowering is mainly regulated by vernalization (VRN1, VRN2, VRN3, and VRN-D4), photoperiod (Ppd-A1, Ppd-B1 Ppd, and Ppd-D1), and earliness per se (eps) genes [39]. Some of these genes showed pleiotropic effects, i.e., Ppd-A1 increased TKW and yield. Similarly, Ppd-B1 was associated with a high kernel number [40]. Eps genes interact with Ppd1 and are associated with spikelet number [41]. In wheat, a homolog of the Arabidopsis early-flowering 3 (ELF3) gene was identified as a candidate gene for Eps-Am1 [42]. Additional genes affecting flowering time in wheat were reviewed by Zhang et al. [39].

Searching Ensembl Plants for wheat orthologs of 204 Arabidopsis thaliana genes related to flowering [43] resulted in a list of 617 wheat genes. Of these, 31 were assigned to some of the SNPs analyzed in this study. Five SNPs were associated with FT or HT, with a negative effect of the ALT allele; these were orthologs of the A. thaliana genes UGT87A2, MFT, FRI, AGL57, TT16, and MAF4.

### 3.2. Plant Height-Related QTLs and Genes

Some Rht loci have been suggested to negatively affect anther extrusion [44,45]. The insertion of Rht genes was crucial for the success of the “green revolution” in wheat. Replacement of some widely used Rht loci with alternative variants more neutral for the flowering process may be recommended in hybrid breeding [46]. For example, the Rht1 and Rht24 genes reduce plant height, but Rht1 simultaneously reduces anther extrusion, whereas Rht24 does not have such adverse effects [30]. All wheats are assumed to be monomorphic for Rht-A1a [47]. Rht24 occurs at relatively high frequencies in European and Chinese wheat cultivars and was mapped to the same region as Rht14, Rht16, and Rht18 [48,49].

QTLs and candidate genes for PH were collected from the catalog of gene symbols for wheat ([47] with updates), and 44 microarray probes reported for PH [50] were mapped

to separate physically linked regions in the Chinese Spring genome V 1.0 (Table S7A). We found eight DArTseq markers overlapping the four regions identified with these microarray probes. Additionally, DArTseq markers were found in the physical regions of QHt.nau-2D, Rht8, Rht13, and Rht22 [51–53].

Information on known genes responsible for wheat PH was used to search for homologous and paralogous genes (with identity to target > 50%). Sequences (TraesCS4A02G271000, TraesCS4B02G043100, and TraesCS4D02G040400) corresponding to gibberellic acid insensitive (GAI) Rht-A1, Rht-B1, and Rht-D1 were mapped to the 4AL, 4BS, and 4DS chromosome arms [54,55]. These GAI loci were highly conserved, and a search with blastn revealed no additional candidate homologous genes.

The remaining Rht genes were sensitive to GA. Two known genes from this group are GA2- $\beta$ -dioxygenase (GA2oxA9) and AP2-D (Q gene), which correspond to Rht18 and Rht23 loci, respectively [50,56]. Two GA2oxA9 homologs and three paralogs were found, and five DArTseq markers coincided with the GA2oxA9 homolog located on the 6B chromosome. Two sequences were used to identify AP-2 homologs. Five AP-2 homologs were found, three of which were located in proximity (<5 Mbp) to 14 DArTseq markers associated with PH (Table S7A).

We identified 224 genes (containing 313 SNPs) assigned to DNA sequences related to PH using the text search in the description of genes by GO terms and Interpro features for the texts “gibberellin,” “auxin,” “cytochrome,” “kaurene,” “kaurenoic,” and “DELLA.” Most detected hits were for “cytochrome,” as in [50]. Of these, only 12 genes (with 14 SNPs) associated with PH showed noticeable negative or positive phenotypic effects. Three of them were orthologs of the *A. thaliana* genes CYP85A1, CYP85A2, CYP84A4, CYP84A1, and CYP734A1.

### 3.3. Spike Traits Related to QTLs and Genes

The positions of significant SNPs for grain weight, number per spike, and TKW were compared with those reported in previous studies (Table S7B). Maintaining a higher KN is an important breeding target for stress-tolerant lines [57]. A total of 142 regions were associated with KN, and 11 overlapped ( $\pm 1$  Mb) with the DArTseq markers reported in our study. Only precise QTLs spanning regions not exceeding 5 Mb were included in the search for common positions with SNP markers identified in our study. Ninety-two markers associated with kernel weight have been described in the literature, and only one marker had a position congruent with the SNP identified in our study. Out of the 123 SNPs or QTL described as responsible for TKW, five SNP markers were identified in previously described regions. TaGW8-B1 is associated with agronomic traits in bread wheat cultivars. The TaGW8-B1a allele increases TKW and spikelet number per spike and provides higher yields than cultivars with the TaGW8-B1b allele [58]. KASP markers for TKW have also been developed [59–62].

### 3.4. SNP Translation Effects

Among the identified polymorphisms, two SNPs had high translational effects. The first identified polymorphism, 1090593 | F | 0-44 | CT, was associated with TKW and located in the gene TraesCS4B02G086500, annotated with the GO term “polygalacturonase activity”. Polygalacturonases are hydrolyzing enzymes implicated in a wide range of plant developmental processes, such as cell elongation, organ abscission, fruit ripening, microspore release, and pollen tube growth [63,64]. Members of two out of the five clades, C and F, are also expressed in grasses during root and seed development [65]. They were detected in the outer pericarp and intermediate layers of grains of the related species *Aegilops tauschii* [66]. Ye et al. [67] found three putative candidate genes for QTL mapped on chromosome 4A in wheat; two (TraesCS4A02G229600 and TraesCS4A02G229700) were orthologous to the *Arabidopsis* gene At2g43860, coding for polygalacturonase. SNP 1090593 | F | 0-44 | CT was localized in the same homologous group but on chromosome 4B and had a high translational level, resulting in a STOP codon instead of a lysine codon. It is predicted that this

substitution resulted in protein shortening (by 40 residues) and an altered C-terminus (the structures of wild-type and altered proteins, Figure S5).

The second SNP (1130302|F|0-44|TG) was located in gene TraesCS6A02G085100 (*Arabidopsis* RPT3 ortholog) with a high translational effect and annotated with the GO term “protein ubiquitination” and was associated with FT. Protein ubiquitination is a sophisticated system of post-translational modification in all eukaryotes and has been demonstrated to play a key role in various plant developmental stages and processes, such as seed dormancy and germination, root growth, flowering time control, self-incompatibility, chloroplast development, and several abiotic stress responses [68]. The SNP 1130302|F|0-44|TG was positioned in the intron and localized in the 5'UTR without any direct impact on the protein sequence. In general, cis-acting elements present in UTRs are essential for post-transcriptional control, including alternative polyadenylation, riboswitching, short-peptide translation, nonsense-mediated decay, and alternative splicing [69]. However, the T/G mutation found in our study may disrupt splicing, possibly resulting in the retention of this intron. Intron retention in the 5'UTR affects gene expression (by alternative splicing, alternative polyadenylation, and protein translation). Such an effect has been shown, *inter alia*, in *Arabidopsis* ZIF2 [70], EF1 $\alpha$ -A3 [71] and rice OsmiR156h [72]. It has been postulated that UTR-related regulation of gene expression helps plants adapt to environmental fluctuations [69]. Flowering time is a major factor in climatic adaptation [73–75].

#### 4. Materials and Methods

##### 4.1. Germplasm Resources

The study was conducted using 509 wheat varieties and breeding lines. A set of 277 European varieties, registered mainly in Germany, Poland, and the United Kingdom, was used. Advanced breeding lines were represented by 232 accessions from the ongoing programs of the Plant Breeding Strzelce (STH) and Poznań Plant Breeding (PHR) companies. The genetic characteristics of the studied resources were provided by Tyrka et al. [28].

##### 4.2. Field Phenotyping

Field experiments were conducted during two vegetation seasons, 2017/18 and 2018/19. Trials were established at four experimental stations of two breeding companies (STH and PHR) located in Strzelce (GPS: 52.3149° N, 19.4025° E), Kończewice (GPS: 53.1848° N, 18.5637° E), Nagradowice (GPS: 52.3178° N, 17.1529° E), and Leszno/Antoniny (GPS: 51.8586° N, 16.5902° E). Two replications of the studied objects were performed at each location. They were distributed within three randomized blocks. Each plot had an area of 1 m<sup>2</sup>. Cultivation conditions in brief: Sowing rate: 250 kernels per m<sup>2</sup>. Fertilization: 45 kg ha<sup>-1</sup> of P<sub>2</sub>O<sub>5</sub> and 80 kg ha<sup>-1</sup> of K<sub>2</sub>O prior sowing followed by 150 kg of N per hectare applied in 2 doses during the vegetation season. Plant protection: only herbicides and insecticides were applied (no treatments against fungi and lodging). The following traits were analyzed in the field experiments:

Time of heading (GS55)—HT (number of days since May 1)

Time of flowering (GS65)—FT (number of days since May 1)

Plant height—PH (average height of plants on a plot in cm)

Number of kernels per spike—KN (assessed on 10 randomly collected spikes in each plot)

Weight of kernels per spike—KW (assessed on 10 randomly collected spikes in each plot; results in grams)

Thousand kernel weight—TKW (average from three random samples per plot, each containing 100 kernels)

To validate the results obtained from the HYBRE experiments, publicly available data from Polish state post-registration trials ([www.coboru.gov.pl](http://www.coboru.gov.pl), accessed on 23 June 2019 [76]) from 2015–2018 were used. Only data on HT, PH, and TKW were available in the official trial reports among the six traits studied in our research.

### 4.3. Genotyping

A detailed description of genotyping and genotypic data processing was provided by Tyrka et al. [28]. Briefly, for each genotype, DNA was isolated from 15–20 bulked 2-week-old seedlings. DNA concentration and purity were determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and DNA quality was assessed using 1.5% agarose gel electrophoresis. The DNA was stored at  $-20^{\circ}\text{C}$  and diluted to a working concentration of  $50\text{ ng}/\mu\text{L}$  for subsequent wheat DArTseq 1.0 genotyping completed by Diversity Arrays Technology (Bruce, Australia). Only selected high-quality data for 13,499 SNPs were taken into account in this research (dominant markers of the Silico-DArT type were omitted).

### 4.4. Data Analysis

Phenotypic data were analyzed using a linear mixed model (LMM) with fixed effects of the year (Y), location (L),  $Y \times L$  interaction, and random effects of genotypes (G),  $G \times Y$  interaction, and  $G \times L$  interaction. Broad-sense heritability was estimated by the method described by Cullis et al. [77]. The additive main effects and multiplicative interaction (AMMI) analysis was performed as described by Gauch [78]. Genome-wide association study (GWAS), using SNP polymorphisms and genotypic means (BLUPs) obtained in individual experiments, was performed by a method that allows for the interaction of genetic effects with the environment, developed by van Eeuwijk et al. [79] and Malosetti et al. [35]. This is based on a linear mixed model with the population structure estimated by eigenanalysis of the kinship matrix (see [28]) and the compound symmetry variance-covariance model used for environmental variation. Marker data used for these analyses were coded as follows: 0, reference (REF) homozygote; 1, heterozygote; and 2, alternative (ALT) homozygote. Standard GWAS considers the additive marker effects, but also non-additive effects might explain an important proportion of the variation in complex traits ([80]). Therefore, GWAS for heterosis effects, estimated as the difference between the mean for heterozygotes and the mean for two homozygotes, was performed based on mean values over all experiments. To do this, marker data coded with 0, 1, and 2 values in the model matrix (used in GWAS for allelic substitution effects) were recoded to values of 0 for homozygotes (of both types) and 1 for heterozygotes [80]. The population structure was represented in the model by the eigenanalysis scores. The effects of marker linkage disequilibrium (LD) clusters were analyzed using mean trait values over all experiments and analysis of covariance, with the genotypic combinations in these clusters as classifying factors and eigenscores as covariates.  $p$ -values for allelic substitution effects and heterosis effects were corrected for multiple testing using the Benjamini–Hochberg (BH) method. Only SNP data with more than five accessions present in both homozygous classes were used during the genome-wide association analyses. An SNP effect was considered statistically significant if the BH-corrected  $p$ -value was lower than 0.05.

Statistical analyses were performed using Genstat for Windows 19th edition [81]. Visualizations of results were performed using Genstat 19 or R software. The annotation of SNP markers with respect to genomic positions, neighboring genes, their Gene Ontology classification, and SNP translation effects [29] used in this study were the same as those used by Tyrka et al. [28].

## 5. Conclusions

In conclusion, by employing an appropriate experimental and statistical approach, we generated a set of SNP markers that can be used in breeding practice to predict the earliness and height of plants to assess whether a given genotype will be particularly good as a maternal (early and short plant) or paternal form (late and tall plant). The estimated heritability of phenological traits and plant height was relatively high; therefore, selection based on the markers indicated during this study may be successful. In contrast, the heritability of yield-related traits in our field trials was relatively low; nevertheless, it remained within the range reported by other authors [82–85].

The main achievements of the presented work are:

1. The successful parallel selection of homozygous parental genotypes (based on traits regulated by additive genes controlling phenology and plant height) and components revealing a high heterosis effect (the choice based on the highest values of spike-yield-related traits revealed by heterozygous genotypes) using GWAS.
2. Demonstrating that the application of clustered markers for the choice of genotypes in multi-feature processes may be more efficient than classic selection based on single marker polymorphism (SNP)
3. Validation of the GWAS results using post-registration trial data.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232315321/s1>.

**Author Contributions:** Conceptualization, M.M., S.S., M.T., M.R.-T. and P.K.; Data curation, M.M.; Formal analysis, M.M. and P.K.; Funding acquisition, M.R.-T.; Methodology, M.M., S.S., M.T., M.R.-T. and P.K.; Resources, P.M., B.Ż., R.M. (Rafał Marcinkowski), U.W.-P., R.M. (Róża Martofel) and M.R.; Visualization, M.M. and P.K.; Writing—original draft, M.M., S.S., M.T., M.R.-T. and P.K.; Writing—review & editing, M.M., S.S., M.T., M.R.-T. and P.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Polish National Centre for Research and Development (project HYBRE, no. BIOSTRATEG3/343665/6/NCBR/2017).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in Supplementary Materials.

**Acknowledgments:** Some computations were performed with the use of resources of Poznań Supercomputing and Networking Center ([www.psnc.pl](http://www.psnc.pl), accessed on 28 November 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

AMMI	Additive main-effects and multiplicative interaction
BLUP	Best linear unbiased prediction
FT	Time of flowering
GBS	Genotyping by Sequencing
GS	Growth stage (BBCH scale)
GWAS	Genome-wide association studies
HYBRE	The project acronym and the name of the consortium
HT	Time of heading
KN	Number of kernels per spike
KW	Weight of kernels per spike
PH	Plant height
PRT	Post-registration trials
SNP	Single Nucleotide Polymorphism
TKW	Thousand kernel weight
VEP	Variant Effect Predictor

## References

1. Poutanen, K.S.; Kårlund, A.O.; Gómez-Gallego, C.; Johansson, D.P.; Scheers, N.M.; Marklinder, I.M.; Eriksen, A.K.; Silventoinen, P.C.; Nordlund, E.; Sozer, N.; et al. Grains—A major source of sustainable protein for health. *Nutr. Rev.* **2022**, *80*, 1648–1663. [[CrossRef](#)] [[PubMed](#)]
2. FAOSTAT. Production domain. In *Crops*; FAO: Rome, Italy, 2019.
3. El Hanafi, S.; Bendaou, N.; Kehel, Z.; Sanchez-Garcia, M.; Tadesse, W. Phenotypic evaluation of elite spring bread wheat genotypes for hybrid potential traits. *Euphytica* **2020**, *216*, 168. [[CrossRef](#)]

4. Briggie, L.W.; Reitz, L.P. *Classification of Triticum Species and of Wheat Varieties Grown in the United States*; U.S. Department of Agriculture: Washington, DC, USA, 1963.
5. Fabrizius, M.A.; Busch, R.H.; Khan, K.; Huckle, L. Genetic diversity and heterosis of spring wheat crosses. *Crop Sci.* **1998**, *38*, 1108–1112. [[CrossRef](#)]
6. Krystkowiak, K.; Adamski, T.; Surma, M.; Kaczmarek, Z. Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (*Triticum aestivum* L.). *Euphytica* **2009**, *165*, 419–434. [[CrossRef](#)]
7. Whitford, R.; Fleury, D.; Reif, J.C.; Garcia, M.; Okada, T.; Korzun, V.; Langridge, P. Hybrid breeding in wheat: Technologies to improve hybrid wheat seed production. *J. Exp. Bot.* **2013**, *64*, 5411–5428. [[CrossRef](#)] [[PubMed](#)]
8. Mühleisen, J.; Piepho, H.-P.; Maurer, H.P.; Longin, C.F.; Reif, J.C. Yield stability of hybrids versus lines in wheat, barley, and triticale. *Theor. Appl. Genet.* **2014**, *127*, 309–316. [[CrossRef](#)] [[PubMed](#)]
9. Creech, C. Making the Case for Certified Wheat Seed. CropWatch. 2017. Available online: <https://cropwatch.unl.edu/2017/making-case-certified-wheat-seed> (accessed on 28 November 2022).
10. Gupta, P.K.; Mir, R.R.; Mohan, A.; Kumar, J. Wheat Genomics: Present Status and Future Prospects. *Int. J. Plant Genom.* **2008**, 896451. [[CrossRef](#)]
11. Wang, S.C.; Wong, D.; Forrest, K.; Allen, A.; Chao, S.; Huang, B.E.; Maccaferri, M.; Salvi, S.; Milner, S.G.; Cattivelli, L.; et al. Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. *Plant Biotechnol.* **2014**, *12*, 787–796. [[CrossRef](#)]
12. Liu, J.D.; He, Z.H.; Rasheed, A.; Wen, W.E.; Yan, J.; Zhang, P.Z.; Wan, Y.X.; Zhang, Y.; Xie, C.J.; Xia, X.C. Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2017**, *17*, 220. [[CrossRef](#)]
13. Liu, J.; Xu, Z.; Fan, X.; Zhou, Q.; Cao, J.; Wang, F.; Ji, G.; Yang, L.; Feng, B.; Wang, T.A. Genome-Wide Association Study of Wheat Spike Related Traits in China. *Front. Plant Sci.* **2018**, *9*, 1584. [[CrossRef](#)]
14. Li, F.J.; Wen, W.E.; He, Z.H.; Liu, J.D.; Jin, H.; Cao, S.H.; Geng, H.W.; Yan, J.; Zhang, P.Z.; Wan, Y.X.; et al. Genome-wide linkage mapping of yield related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor. Appl. Genet.* **2018**, *131*, 1903–1924. [[CrossRef](#)] [[PubMed](#)]
15. Muqaddasi, Q.H.; Lohwasser, U.; Nagel, M.; Börner, A.; Pillen, K.; Röder, M.S. Genome-Wide Association Mapping of Anther Extrusion in Hexaploid Spring Wheat. *PLoS ONE* **2016**, *11*, e0155494. [[CrossRef](#)]
16. Adhikari, A.; Basnet, B.R.; Crossa, J.; Dreisigacker, S.; Camarillo, F.; Bhati, P.K.; Jarquin, D.; Manes, Y.; Ibrahim, A.M.H. Genome-Wide Association Mapping and Genomic Prediction of Anther Extrusion in CIMMYT Hybrid Wheat Breeding Program via Modeling Pedigree, Genomic Relationship, and Interaction With the Environment. *Front. Genet.* **2020**, *11*, 586687. [[CrossRef](#)] [[PubMed](#)]
17. Gahlaut, V.; Jaiswal, V.; Singh, S.; Balyan, H.S.; Gupta, P.K. Multi-Locus Genome Wide Association Mapping for Yield and Its Contributing Traits in Hexaploid Wheat under Different Water Regimes. *Sci. Rep.* **2019**, *9*, 19486. [[CrossRef](#)] [[PubMed](#)]
18. Jamil, M.; Ali, A.; Gul, A.; Ghafoor, A.; Napar, A.A.; Ibrahim, A.M.H.; Naveed, N.H.; Yasin, N.A.; Mujeeb-Kazi, A. Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol.* **2019**, *19*, 149. [[CrossRef](#)]
19. Li, F.; Wen, W.; Liu, J.; Zhang, Y.; Cao, S.; He, Z.; Rasheed, A.; Jin, H.; Zhang, C.; Yan, J.; et al. Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* **2019**, *19*, 168. [[CrossRef](#)]
20. Tsai, H.-Y.; Janss, L.L.; Andersen, J.R.; Orabi, J.; Jensen, J.D.; Jahoor, A.; Jensen, J. Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Sci. Rep.* **2020**, *10*, 3347. [[CrossRef](#)]
21. De Vries, A.P. Flowering biology of wheat, particularly in view of hybrid seed production—A review. *Euphytica* **1971**, *20*, 152–170. [[CrossRef](#)]
22. Longin, F.; Mühleisen, J.; Maurer, H.P.; Zhang, H.; Gowda, M.; Reif, J. Hybrid breeding in autogamous cereals. *Theor. Appl. Genet.* **2012**, *125*, 1087–1096. [[CrossRef](#)]
23. Langer, S.M.; Longin, C.F.H.; Würschum, T. Phenotypic evaluation of floral and flowering traits with relevance for hybrid breeding in wheat (*Triticum aestivum* L.). *Plant Breed.* **2014**, *133*, 433–441. [[CrossRef](#)]
24. Okada, T.; Ridma, J.E.A.; Jayasinghe, M.; Nansamba, M.; Baes, M.; Warner, P.; Kouidri, A.; Correia, D.; Nguyen, V.; Whitford, R.; et al. Unfertilized ovary pushes wheat flower for cross-pollination. *J. Exp. Bot.* **2018**, *69*, 399–412. [[CrossRef](#)] [[PubMed](#)]
25. Selva, C.; Riboni, M.; Baumann, U.; Würschum, T.; Whitford, R.; Tucker, M.R. Hybrid breeding in wheat: How shaping floral biology can offer new perspectives. *Funct. Plant Biol.* **2020**, *47*, 675–694. [[CrossRef](#)] [[PubMed](#)]
26. Zajączkowska, U.; Denisow, B.; Łotocka, B.; Dołkin-Lewko, A.; Rakoczy-Trojanowska, M. Spikelet movements, anther extrusion and pollen production in wheat cultivars with contrasting tendencies to cleistogamy. *BMC Plant Biol.* **2021**, *21*, 136. [[CrossRef](#)]
27. Hayes, J.D. Hybrid Wheat—Results and Problems, by A. A. Pickett. 259 pp. Berlin, Hamburg: Paul Parey. DM 96.00 (paperback). ISBN 3 489 53510 3. *J. Agric. Sci.* **1993**, *121*, 294. [[CrossRef](#)]
28. Tyrka, M.; Mokrzycka, M.; Bakera, B.; Tyrka, D.; Szeliga, M.; Stojalowski, S.; Matysik, P.; Rokicki, M.; Rakoczy-Trojanowska, M.; Krajewski, P. Evaluation of genetic structure in European wheat cultivars and advanced breeding lines using high-density genotyping-by sequencing approach. *BMC Genom.* **2021**, *22*, 81. [[CrossRef](#)] [[PubMed](#)]
29. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [[CrossRef](#)]

30. Gupta, D.R.; Avila, C.S.R.; Win, J.; Soanes, D.M.; Ryder, L.S.; Croll, D.; Bhattacharjee, P.; Hossain, M.S.; Mahmud, N.U.; Meheub, M.S.; et al. Cautionary Notes on Use of the MoT3 Diagnostic Assay for Magnaporthe oryzae Wheat and Rice Blast Isolates. *Phytopathology* **2019**, *109*, 504–508. [CrossRef] [PubMed]
31. Waines, J.G.; Hegde, S.G. Intraspecific Gene Flow in Bread Wheat as Affected by Reproductive Biology and Pollination Ecology of Wheat Flowers. *Crop Sci.* **2003**, *43*, 451–463. [CrossRef]
32. Racedo, J.; Gutiérrez, L.; Perera, M.F.; Ostengo, S.; Pardo, E.M.; Cuenya, M.I.; Welin, B.; Castagnaro, A.P. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol.* **2016**, *16*, 142. [CrossRef]
33. Mwadzingeni, L.; Shimelis, H.; Rees, D.J.G.; Tsilo, T.J. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. *PLoS ONE* **2017**, *12*, e0171692. [CrossRef]
34. Mourad, A.M.I.; Sallam, A.; Belamkar, V.; Wegulo, S.; Bowden, R.; Jin, Y.; Mahdy, E.; Bakheit, B.; El-Wafaa, A.A.; Poland, J.; et al. Genome-Wide Association Study for Identification and Validation of Novel SNP Markers for Sr6 Stem Rust Resistance Gene in Bread Wheat. *Front. Plant Sci.* **2018**, *9*, 380. [CrossRef] [PubMed]
35. Malosetti, M.; Ribaut, J.M.; van Eeuwijk, F.A. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Front. Physiol.* **2013**, *4*, 44. [CrossRef] [PubMed]
36. Denisow, B.; Masierowska, M.; Winiarczyk, K.; Rakoczy-Trojanowska, M. The pollen dispersal ability for cross-pollination in winter wheat (*Triticum aestivum* L.) is related to anther extrusion capability rather than to pollen output. *S. Afr. J. Bot.* **2022**, *148*, 283–292. [CrossRef]
37. Allan, R.E. Wheat. In *Hybridization of Crop Plants*; Agronomy Books. 3; Fehr, W.R., Hadley, H.H., Eds.; Iowa State University: Ames, IA, USA, 1980; Available online: [https://lib.dr.iastate.edu/agron\\_books/3](https://lib.dr.iastate.edu/agron_books/3) (accessed on 28 November 2022).
38. Wilkinson, P.A.; Winfield, M.O.; Barker, G.L.A.; Allen, A.M.; Burrige, A.; Coghill, J.A.; Burrige, A.; Edwards, K.J. CerealsDB 2.0: An integrated resource for plant breeders and scientists. *BMC Bioinform.* **2013**, *13*, 219. [CrossRef]
39. Zhang, J.; She, M.; Yang, R.; Jiang, Y.; Qin, Y.; Zhai, S.; Balotf, S.; Zhao, Y.; Anwar, M.; Alhabbar, Z.; et al. Yield-Related QTL Clusters and the Potential Candidate Genes in Two Wheat DH Populations. *Int. J. Mol. Sci.* **2021**, *22*, 11934. [CrossRef]
40. Arjona, J.; Royo, C.; Dreisigacker, S.; Ammar, K.; Villegas, D. Effect of Ppd-A1 and Ppd-B1 Allelic Variants on Grain Number and Thousand Kernel Weight of Durum Wheat and Their Impact on Final Grain Yield. *Front. Plant Sci.* **2018**, *9*, 888. [CrossRef]
41. Zikhali, M.; Griffiths, S. The effect of Earliness per se (Eps) genes on flowering time in bread wheat. In *Advances in Wheat Genetics: From Genome to Field, Proceedings of the 12th International Wheat Genetics Symposium, Yokohama, Japan, 8–14 September 2013*; Yasunari, O., Shigeo, T., Hirokazu, H., Eds.; Springer: Tokyo, Japan, 2015; pp. 339–345. [CrossRef]
42. Wang, J.; Wen, W.; Hanif, M.; Xia, X.; Wang, H.; Liu, S.; Liu, J.; Yang, L.; Cao, S.; He, Z. TaELF3-1DL, a homolog of ELF3, is associated with heading date in bread wheat. *Mol. Breed.* **2016**, *36*, 161. [CrossRef]
43. Peng, F.Y.; Hu, Z.; Yang, R.-C. Genome-Wide Comparative Analysis of Flowering-Related Genes in Arabidopsis, Wheat, and Barley. *Int. J. Plant Genom.* **2015**, 874361. [CrossRef]
44. Lu, Q.; Lillemo, M.; Skinnis, H.; He, X.; Shi, J.; Ji, F.; Dong, Y.; Bjørnstad, A. Anther extrusion and plant height are associated with Type I resistance to Fusarium head blight in bread wheat line ‘Shanghai-3/Catbird’. *Theor. Appl. Genet.* **2013**, *126*, 317–334. [CrossRef]
45. Boeven, P.H.G.; Longin, C.F.H.; Leiser, W.L.; Kollers, S.; Ebmeyer, E.; Würschum, T. Genetic architecture of male floral traits required for hybrid wheat breeding. *Theor. Appl. Genet.* **2016**, *129*, 2343–2357. [CrossRef]
46. Würschum, T.; Liu, G.; Boeven, P.H.G.; Longin, C.F.H.; Mirdita, V.; Kazman, E.; Zhao, Y.; Reif, J.C. Exploiting the Rht portfolio for hybrid wheat breeding. *Theor. Appl. Genet.* **2018**, *131*, 1433–1442. [CrossRef] [PubMed]
47. McIntosh, R.A.; Yamazaki, Y.; Devos, K.M.; Dubcovsky, J.; Rogers, W.J.; Appels, R. Catalogue of gene symbols for wheat. In *Proceedings of the 10th International Wheat Genetics Symposium, Paestrum, Italy, 1–6 September 2003*; Pogna, N.E., Romano, N., Pogna, E.A., Galterio, G., Eds.; Instituto Sperimentale per la Cerealicoltura: Rome, Italy, 2003; pp. 1–34.
48. Tian, X.; Wen, W.; Xie, L.; Fu, L.; Xu, D.; Fu, C.; Wang, D.; Chen, X.; Xia, X.; Chen, Q.; et al. Molecular Mapping of Reduced Plant Height Gene Rht24 in Bread Wheat. *Front. Plant Sci.* **2017**, *8*, 998. [CrossRef]
49. Würschum, T.; Langer, S.M.; Longin, C.F.H.; Tucker, M.R.; Leiser, W.L. A modern Green Revolution gene for reduced height in wheat. *Plant J.* **2017**, *92*, 892–903. [CrossRef] [PubMed]
50. Zanke, C.D.; Ling, J.; Plieske, J.; Kollers, S.; Ebmeyer, E.; Korzun, V.; Argillier, O.; Stiewe, G.; Hinze, M.; Neumann, K.; et al. Whole Genome Association Mapping of Plant Height in Winter Wheat (*Triticum aestivum* L.). *PLoS ONE* **2014**, *9*, e113287. [CrossRef]
51. Korzun, V.; Röder, M.; Ganai, M.W.; Worland, A.J.; Law, C.N. Genetic analysis of the dwarfing gene (Rht8) in wheat. Part I. Molecular mapping of Rht8 on the short arm of chromosome 2D of bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **1998**, *96*, 1104–1109. [CrossRef]
52. Ellis, M.H.; Rebetzke, G.J.; Azanza, F.; Richards, R.A.; Spielmeyer, W. Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. *Theor. Appl. Genet.* **2005**, *111*, 423–430. [CrossRef]
53. Peng, Z.S.; Li, X.; Yang, Z.J.; Liao, M.L. A new reduced height gene found in the tetraploid semi-dwarf wheat landrace Aiganfanmai. *Genet. Mol. Res.* **2011**, *10*, 2349–2357. [CrossRef]
54. Ellis, H.; Spielmeyer, W.; Gale, R.; Rebetzke, J.; Richards, A. “Perfect” markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. *Theor. Appl. Genet.* **2002**, *105*, 1038–1042. [CrossRef]
55. Pearce, S.; Saville, R.; Vaughan, S.P.; Chandler, P.M.; Wilhelm, E.P.; Sparks, C.A.; Al-Kaff, N.; Korolev, A.; Boulton, M.I.; Phillips, A.L.; et al. Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. *Plant Physiol.* **2011**, *157*, 1820–1831. [CrossRef]

56. Xu, T.; Bian, N.; Wen, M.; Xiao, J.; Yuan, C.; Cao, A.; Zhang, S.; Wang, X.; Wang, H. Characterization of a common wheat (*Triticum aestivum* L.) high-tillering dwarf mutant. *Theor. Appl. Genet.* **2017**, *130*, 483–494. [CrossRef]
57. Alemu, A.; Suliman, S.; Hagra, A.; Thabet, S.; Al-Abdallad, A.; Abdelmula, A.A.; Tadesse, W. Multi-model genome-wide association and genomic prediction analysis of 16 agronomic, physiological and quality related traits in ICARDA spring wheat. *Euphytica* **2021**, *217*, 205. [CrossRef]
58. Yan, X.; Zhao, L.; Ren, Y.; Dong, Z.; Cui, D.; Chen, F. Genome-wide association study revealed that the TaGW8 gene was associated with kernel size in Chinese bread wheat. *Sci. Rep.* **2019**, *9*, 2702. [CrossRef] [PubMed]
59. Yang, J.; Zhou, Y.; Wu, Q.; Chen, Y.; Zhang, P.; Zhang, Y.; Cao, T. Molecular characterization of a novel TaGL3–5A allele and its association with grain length in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2019**, *132*, 1799–1814. [CrossRef] [PubMed]
60. Yang, J.; Zhou, Y.; Zhang, Y.; Hu, W.; Wu, Q.; Chen, Y.; Zhao, H. Cloning, characterization of TaGS3 and identification of allelic variation associated with kernel traits in wheat (*Triticum aestivum* L.). *BMC Genet.* **2019**, *20*, 98. [CrossRef] [PubMed]
61. Yang, L.; Zhao, D.; Meng, Z.; Xu, K.; Yan, J.; Xia, X.; Zhang, Y. QTL mapping for grain yield-related traits in bread wheat via SNP-based selective genotyping. *Theor. Appl. Genet.* **2019**, *133*, 857–872. [CrossRef] [PubMed]
62. Daba, S.D.; Tyagi, P.; Brown-Guedira, G.; Mohammadi, M. Genome-wide association studies to identify loci and candidate genes controlling kernel weight and length in a historical United States wheat population. *Front. Plant Sci.* **2018**, *9*, 1045. [CrossRef]
63. Hadfield, K.A.; Bennett, A.B. Polygalacturonases: Many genes in search of a function. *Plant Physiol.* **1998**, *117*, 337–343. [CrossRef]
64. Babu, Y.; Bayer, M. Plant Polygalacturonases Involved in Cell Elongation and Separation-The Same but Different? *Plants* **2014**, *3*, 613–623. [CrossRef]
65. Yang, Y.; Yu, Y.; Liang, Y.; Anderson, C.T.; Cao, J. A Profusion of Molecular Scissors for Pectins: Classification, Expression, and Functions of Plant Polygalacturonases. *Front. Plant Sci.* **2018**, *9*, 1208. [CrossRef]
66. Chateigner-Boutin, A.L.; Suliman, M.; Bouchet, B.; Alvarado, C.; Lollier, V.; Rogniaux, H.; Guillon, F.; Larré, C. Endomembrane proteomics reveals putative enzymes involved in cell wall metabolism in wheat grain outer layers. *J. Exp. Bot.* **2015**, *66*, 2649–2658. [CrossRef]
67. Ye, X.; Li, J.; Cheng, Y.; Yao, F.; Long, L.; Wang, Y.; Wu, Y.; Li, J.; Wang, J.; Jiang, Q.; et al. Genome-wide association study reveals new loci for yield-related traits in Sichuan wheat germplasm under stripe rust stress. *BMC Genom.* **2019**, *20*, 640. [CrossRef] [PubMed]
68. Shu, K.; Yang, W. E3 Ubiquitin Ligases: Ubiquitous Actors in Plant Development and Abiotic Stress Responses. *Plant Cell Physiol.* **2017**, *58*, 1461–1476. [CrossRef] [PubMed]
69. Srivastava, A.K.; Lu, Y.; Zinta, G.; Lang, Z.; Zhu, J.K. UTR-Dependent Control of Gene Expression in Plants. *Trends Plant Sci.* **2018**, *23*, 248–259. [CrossRef] [PubMed]
70. Remy, E.; Cabrito, T.R.; Batista, R.A.; Hussein, M.A.M.; Teixeira, M.C.; Athanasiadis, A.; Sá-Correia, I.; Duque, P. Intron Retention in the 5'UTR of the Novel ZIF2 Transporter Enhances Translation to Promote Zinc Tolerance in Arabidopsis. *PLoS Genet.* **2014**, *10*, e1004375. [CrossRef]
71. Chung, B.Y.; Simons, C.; Firth, A.E.; Brown, C.M.; Hellens, R.P. Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. *BMC Genom.* **2006**, *7*, 120. [CrossRef]
72. Zhao, M.; Liu, B.; Wu, K.; Ye, Y.; Huang, S.; Wang, S.; Wang, Y.; Han, R.; Liu, Q.; Fu, X.; et al. Regulation of OsmiR156h through Alternative Polyadenylation Improves Grain Yield in Rice. *PLoS ONE* **2015**, *10*, e0126154. [CrossRef] [PubMed]
73. Jung, C.; Muller, A.E. Flowering time control and applications in plant breeding. *Trends Plant Sci.* **2009**, *14*, 563–573. [CrossRef]
74. Gaudinier, A.; Blackman, B.K. Evolutionary processes from the perspective of flowering time diversity. *New Phytol.* **2020**, *225*, 1883–1898. [CrossRef]
75. Wu, C.-C.; Wei, F.-J.; Chiou, W.-Y.; Tsai, Y.-C.; Wu, H.-P.; Gotarkar, D.; Wei, Z.H.; Lai, M.H.; Hsing, Y.I.C. Studies of rice Hd1 haplotypes worldwide reveal adaptation of flowering time to different environments. *PLoS ONE* **2020**, *15*, e0239028. [CrossRef]
76. Research Centre for Cultivar Testing. Available online: <http://www.coboru.pl> (accessed on 23 June 2019).
77. Cullis, B.R.; Smith, A.B.; Coombes, N.E. On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* **2006**, *11*, 381–393. [CrossRef]
78. Gauch, H.G. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*; Elsevier: Amsterdam, The Netherlands, 1992.
79. van Eeuwijk, F.A.; Bink, M.C.A.M.; Chenu, K.; Chapman, S.C. Detection and use of QTL for complex traits in multiple environments. *Curr. Opin. Plant Biol.* **2010**, *13*, 193–205. [CrossRef] [PubMed]
80. Vidotti, M.S.; Lyra, D.H.; Morosini, J.S.; Granato, Í.S.C.; Quecine, M.C.; Azevedo, J.L.D.; Fritsche-Neto, R. Additive and heterozygous (dis)advantage GWAS models reveal candidate genes involved in the genotypic variation of maize hybrids to *Azospirillum brasilense*. *PLoS ONE* **2019**, *14*, e0222788. [CrossRef] [PubMed]
81. VSN International. Genstat for Windows. 19th ed. Hemel Hempstead. Web Page. 2017. Available online: [Genstat.co.uk](http://Genstat.co.uk) (accessed on 28 November 2022).
82. Kashif, M.; Khaliq, I. Heritability, correlation and path coefficient analysis for some metric traits in wheat. *Int. J. Agric. Biol.* **2004**, *6*, 138–142.
83. Nukasani, V.; Potdukhe, N.R.; Bharad, S.; Deshmukh, S.; Shinde, S.M. Genetic variability, correlation and path analysis in wheat. *J. Wheat Res.* **2013**, *5*, 48–51.

84. Guo, Z.; Zhao, Y.; Röder, M.S.; Reif, J.C.; Ganai, M.W.; Chen, D.; Schnurbusch, T. Manipulation and prediction of spike morphology traits for the improvement of grain yield in wheat. *Sci. Rep.* **2018**, *8*, 14435. [[CrossRef](#)]
85. Mangini, G.; Gadaleta, A.; Colasuonno, P.; Marcotuli, I.; Signorile, A.M.; Simeone, R.; De Vita, P.; Mastrangelo, A.M.; Laidò, G.; Pecchioni, N.; et al. Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PLoS ONE* **2018**, *13*, e0190162. [[CrossRef](#)]