Article

Identification of markers associated with Wheat dwarf virus (WDV) tolerance/resistance in barley (Hordeum vulgare ssp. vulgare) using Genome wide association studies

Behnaz Soleimani[1,*], Heike Lehnert[2], Sarah Trebing[1], Antje Habekuß[1], Frank Ordon[1], Andreas Stahl[1] and Torsten Will[1]

[1] Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Julius Kuehn Institute, Quedlinburg, Germany

[2] Institute for Biosafety in Plant Biotechnology, Julius Kuehn Institute, Quedlinburg, Germany

* Correspondence: behnaz.soleimani@julius-kuehn.de

Abstract:

Wheat dwarf virus (WDV) causes an important vector transmitted virus disease, which leads to significant yield losses in barley production. Due to the fact, that at the moment no plant protection products are approved to combat the vector *Psammotettix alienus* and this disease cannot be controlled by chemical means, therefore the use of WDV resistant or tolerant genotypes is the most efficient method to control and reduce negative effects of WDV on barley growth and production. In this study, a set of 480 barley genotypes were screened to identify genotypic differences in response to WDV and five traits were assessed under infected and non-infected conditions. In total, 32 genotypes showed resistance or tolerance against WDV. Subsequently, the phenotypic data of 191 out of 480 genotypes combined with 34,408 single nucleotide polymorphisms (SNPs) in a genome wide association study to identify quantitative trait loci (QTLs) and markers linked to resistance/tolerance to WDV, using Tassel, GAPIT, and FARM CPU. In total, 1, 3, 2, 2 and 1 significantly associated markers on chromosomes 3H, 4H, 5H and 7H identified by all three methods for ELISA-60, relative performance of total grain weight, plant height, number of ears per plant and thousand grain weight, respectively.

Keywords: barley, *Hordeum vulgare* ssp. *vulgare*, *Wheat dwarf virus* (WDV), leafhoppers, resistance, tolerance, single nucleotide polymorphisms (SNPs), marker, QTL

1. Introduction

Adapting barley cultivars to a changing production environment is a contemporary task of barley breeding. Barley is ranked the fourth most important crop for food and feed, worldwide [1] and its cultivation is threatened by abiotic and biotic stresses. Changes in climate conditions especially associated with increasing temperatures will promote the occurrence and development of insect and virus populations [2]. In detail, it is described that longer periods of high temperatures during autumn and winter lead to an increased occurrence of insect transmitted virus disease, i.e. the aphid-transmitted *Barley yellow dwarf virus* (BYDV) and the leafhopper-transmitted *Wheat dwarf virus* (WDV) [2]. Wheat dwarf virus (WDV) is known as an important cereal pathogen [3], which is transmitted by the leafhopper *Psammotettis alienus* (*Cicadelliae* family). WDV belongs to the family *Geminiviridae* and the genus *Mastrevirus*. WDV has a monopartite genome (genome size 2.7 kb) with single-stranded circular DNA [4]. The virus causes severe symptoms in barley like dwarfing, tufting, streaks and leaf chlorosis, reduced spike number and yield losses [3, 5, 6]. Negative effects of virus infection on yield were described for nearly whole Europe as well as for parts of Africa and Asia [3]. The

presence of WDV in Europe was firstly reported by Vacke [7] in the former Czechoslovakia. Later, the occurrence of the virus was also reported from other European countries, i.e. Sweden, Hungary, France and Germany [8], and some parts of Africa and Asia [3]. WDV is able to infect different species of the *Poaceae* family such as barley, wheat, oat, rye, maize and many wild grasses. Therefore, it might be considered as a grass generalist pathogen [3]. Due to the lack of insecticides, but also with regard to the goal of reducing pesticide application according to the farm to fork strategy within the European Green Deal, direct control of *P. alienus* with insecticides is currently not possible and will most likely not be feasible in future. Therefore, identifying virus resistant or tolerant barley genotypes is the best way to avoid negative effects of WDV in the future.

Today, Next generation sequencing (NGS) or array-based technologies enable genotyping of diverse genotype collections in a short time and with high accuracy [9]. High dense marker sets made it possible to identify marker–trait associations (MTAs) and quantitative trait loci (QTL) by mapping studies or genome wide association studies (GWAS) [10]. Several programs are available to conduct GWAS, e.g. TASSEL [11], PLINK [12], R (GAPIT [13]) and FARMCPU [14]) software. Several QTL regions associated with different traits were already identified by GWAS in barley for yield, seed quality and disease-related traits [15], e.g., spot blotch resistance [16, 17], abiotic stresses such as drought stress [18-20]. However, until now, no QTL regions involved in tolerance or resistance to WDV were identified in barley, but recently for wheat [21]. Identification of QTL regions and development of diagnostic markers associated with tolerance or resistance to WDV are important and will be helpful for future barley breeding programs. Therefore, the present study focused on the identification of QTL regions, associated with tolerance or resistance to WDV in barley. To achieve this, we tested a diverse collection of winter barley genotypes (the primary gene pool of barley) for WDV tolerance and conducted a GWAS to identify quantitative trait loci (QTL) for WDV tolerance.

2. Materials and Methods

2.1. Plant material

A set of 480 barley genotypes was selected for the study. Seeds of all genotypes were kindly provided by the gene banks of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben and the NATIONAL PLANT GERMPLASM SYSTEM, United States. Selected genotypes originated from four different continents and 22 countries (Table S1).The majority under investigation originated from the Fertile Crescent (Middle East) area, which is considered as a diversity center for barley. The two barley genotypes "Rubina" and "Post" were used as susceptible and tolerant control genotypes in all evaluations.

2.1.1. Phenotyping

The tolerance/resistance test of the entire set of 480 barley genotypes was performed in two successive years. A subset of 240 genotypes was sown in September 2016 and the remaining genotypes were tested in the following year in three gauze houses respectively (Table S1). In addition, a set of 50 promising tolerant/resistant genotypes from the first year were tested in the second year, as well. Genotypes which indicated extinction values below the cut-off [21] were excluded from further analysis and a set of 250 out of 480 tested genotypes were genotyped by 50K the iSelect chip [22]. Ultimately, based upon the availability of phenotypic and genotypic data, a subset of 191 barley genotypes was considered for conducting GWAS and further analysis.

The resistance tests were carried out in three neighboring gauze houses located on the field of the experimental station of the Julius Kuehn Institute in Quedlinburg, Germany   (51°46′20.7″N 11°08′46.5″E). The collection of 480 barley genotypes was phenotyped under inoculated (I, I-variant) and non-inoculated (control, C-variant) conditions. Ten to 15 seeds per genotype and variant were sown in a row. The WDV inoculation was conducted at BBCH 11-12, i.e. the one to two leaf stage. To increase the infection pressure, one WDV-infected barley plant was placed in a short distance between each row of the inoculated variant [2]. Before the infestation of plants by leafhoppers, all plants

were covered by additional gauze tunnels. The virus bearing leafhoppers were distributed at a stocking density of approx. 1 insect/plant [2]. Insecticides were applied four weeks after the inoculation and subsequently gauze tunnels were removed.

First phenotyping was performed at BBCH 23-30. Leaf samples of all plants grown under infected conditions were taken and 50 mg of leaf material were used to detect a WDV infection by using double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA). The second phenotyping was carried out visually (symptom scoring scale 1 to 9; 1= free of symptoms, 9= dead plant, (Pfrieme et al., 2022)) at BBCH 59 [2]. The scoring values were used to discriminate resistant, tolerant and susceptible genotypes, as well. For instance, genotypes with scoring value 1 (Pfrieme et al., 2022) and low WDV infection rate were considered as resistant. Genotypes with scoring value of 2-4 (Pfrieme et al., 2022) and low infection rate were considered as tolerant. Genotypes with scoring value 5 to 9 (Pfrieme et al., 2022) and high infection rate were grouped as susceptible genotypes. At the end of each experiment at BBCH 99   total grain weight (ToGW), plant height (HEI), number of ears per plant (NEP) and thousand grain weight (TGW) (Table 1) were measured under I-variant and C-variant.

Table 1. List of evaluated traits under two different variants.

Trait	Abbreviation	Method of measurement	Unit
ELISA-60		double antibody sandwich enzyme-linked immunosorbent assay	
Total grain weight	ToGW	Weight all harvested seeds per plant	g
Plant height	HEI	Measure plant length from basis to leaf tip	cm
Number of ears per plant	NEP	Count number of ears after harvesting	
Thousand grain weight	TGW	Weigh 1000 grain after threshing	g

The relative performance was determined for each trait by applying the following formula:

$$\text{Relative performace} = \frac{Gi}{Gc}$$

where Gi and Gc are the mean trait performances of a barley genotype under infected conditions or non-infected conditions, respectively. The relative performance of each trait was used as phenotypic input for GWAS.

2.2. Statistical analysis

The analysis of phenotypic data was conducted by using the SAS Enterprise Guide 7.1 (SAS Institute Inc., Cary, NC, USA). Quality check of raw data was carried out to exclude outliers, i.e. a value lower or higher than 2 standard deviations. The arithmetic mean value per genotype was estimated for each trait after removing outliers. The Shapiro-Wilk test to evaluate normality of data was performed. The procedure PROC MIXED was used for analysis of variance (ANOVA). Two mixed models were calculated: model 1 was applied to total grain weight, plant height, number of ears per plant and thousand grain weight and model 2 was applied to ELISA values, respectively.

Model 1: $Y_{ijklm} = \mu + T_i + G_j + T_i \times G_j + Y_k + Y_k \times H_l(row_m) + e_{ijklm}$
Model 2: $Y_{jklm} = \mu + G_j + Y_k + Y_k \times H_l(row_m) + e_{jklm}$

Y_{ijklm} is the phenotypic value of the jth genotype in the ith treamtent in k^{th} year in l^{th} and m^{th}, Y_{jklm} is the phenotypic value of the jth genotype in the kth year in l^{th} and m^{th}. μ is the general mean, T_i is the fixed effect of the ith treatment, G_j is the fixed effect of the jth genotype, $T_i \times G_j$ is the fixed interaction

effect between the ith treatment and j^{th} genotype, Y_k and Y_k x $H_l(row_m)$ is the random effect of the kth year and the lth gauze houses, nested in the mth row and e is the random error term.

2.3. Genome wide association study (GWAS)

In total, 191 out of 480 genotypes were used for GWAS analysis, due to the availability of phenotypic and genotypic data for all five measured traits. Genomic DNA was extracted using a modified CTAB method based on Doyle and Doyle [23]. Genotyping was carried out by TraitGenetics (SGS Institute Fresenius GmbH, Gatersleben, Germany) using a 50K iSelect chip (Illumina Inc., San Diego, USA), which resulted in genome wide marker data of 44,040 single nucleotide polymorphisms (SNPs). The reference genome of the barley cultivar Morex (Morex V2) [24] was used   for mapping flanking marker sequences against the reference genome. All mapped markers were filtered for ≥ 30% missing value and monomorphic markers. In a third step, SNP imputation was carried out using the software package Beagle version 4.1 [25, 26]. The imputed marker data set was filtered for minor allele frequency (MAF) ≥ 5% and heterozygosity ≤12.5%. Finally, a set of 34,408 markers was used for further analyses.

A set of 3,117 highly informative markers was used to calculate genetic distance for Kinship matrix and population structure. This set of markers consisted of independent markers in linkage equilibrium (LE) and was selected by using the Plink software [12].

Rogers' distances (RD) were estimated for pairwise genotype–genotype combinations [27] and transformed in a similarity matrix. Subsequently, this matrix was used as kinship matrix. Population structure was determined by using Bayesian cluster analysis implemented in the STRUCTURE software package version 2.3.4 [28] and principal coordinate analysis (PCoA) implemented in the DARwin 6 software [29]. The number of clusters (k) was set from 1 to 10. Structure was started with 10 independent runs for each k. The burn in time and Markov Chain Monte Carlo (MCMC) iterations were set to 100,000. The optimal number of subpopulations was determined by using the Evanno method (ΔK method) implemented in the STRUCTURE HARVESTER software (http://taylor0.biology.ucla.edu/structureHarvester [30]).

In total, 34,408 SNP markers and phenotypic data for five traits were used to perform GWAS. Three different programs, i.e., Tassel, GAPIT and FARM CPU were used independently. The following models were used to identify significant marker trait associations: 1) Mixed Linear Model (MLM) in TASSEL 5.0 [11], which included K matrix and a Q-matrix as correction for population structure, 2) Compressed Mixed Linear Model (CMLM) in GAPIT [13], which included a K-matrix and Q-matrix as correction for population structure and 3) Fixed and Random Model Circulating Probability Unification (FARMCPU [14]), which included a Q-matrix based. In addition, MLM and CMLM models were applied only with K matrix.

Identified significant markers with overlaps in Tassel, GAPIT and FARM CPU were assigned to QTL regions based on their chromosomal position and the linkage disequilibrium (LD) decay. LD decay was estimated by using the software package R [31], packages "genetics" and "LDheatmap" [32, 33]. The LD was calculated as squared allelic correlation (r^2) between all pairs of markers within a chromosome. The genetic distances between markers in base pairs were plotted against the estimated r^2. The r^2 values were set to 0.2 [34]. To estimate the LD decay, a locally weighted polynomial regression (LOESS) curve was fitted [35]. Finally, the intersection of the LOESS curve and the critical r^2 value were used to determine the LD decay [35, 36]. The LD decay was separately calculated for each single chromosome and across all seven barley chromosomes. The threshold of LOD ≥3 (-log10 of P value) was used to determine the significant associations.

3. Results

3.1. Phenotypic data

Barley genotypes showed different phenotypic reactions in response to a WDV infection (I-variant). Trait performances of all 480 tested genotypes are shown in Table S2. In the following, trait

performances and ANOVA result of studied traits are presented for the subset of 191 genotypes. The lower mean value under I-variant was observed for all four tested traits (ToGW, HEI, NEP and TGW) (Table 2, Figure 1).

Table 2. Trait performance of 191 investigated barley genotypes.

Trait	Treatment[a]	N[b]	Mean[c]	Minimum[d]	Maximum[d]	Sd[e]	CV[f]
ELISA-60	I-variant	1866.0	0.3	-0.03	1.84	0.6	179.7
Total grain weight	I-variant	209	11.3	0	287.4	31.7	280.7
	C-variant	229	156.3	2.2	781.7	102.6	65.7
Plant height	I-variant	520	62.9	1	159	40.4	64.2
	C-variant	1163	117.5	49	224	20.3	17.2
Number of ears per plant	I-variant	524	6.9	0	52	8.2	118.3
	C-variant	1160	18.4	1	117	12.1	65.8
Thousand grain weight	I-variant	208	25.5	5	48.3	8.7	34.1
	C-variant	229	43.5	25.37	64.8	7.8	17.9

[a] Treatment: infected (I-variant) and control (C- Variant)

[b] Number of observations

[c] Mean value

[d] Maximum and Minimum

[e] Standard deviation

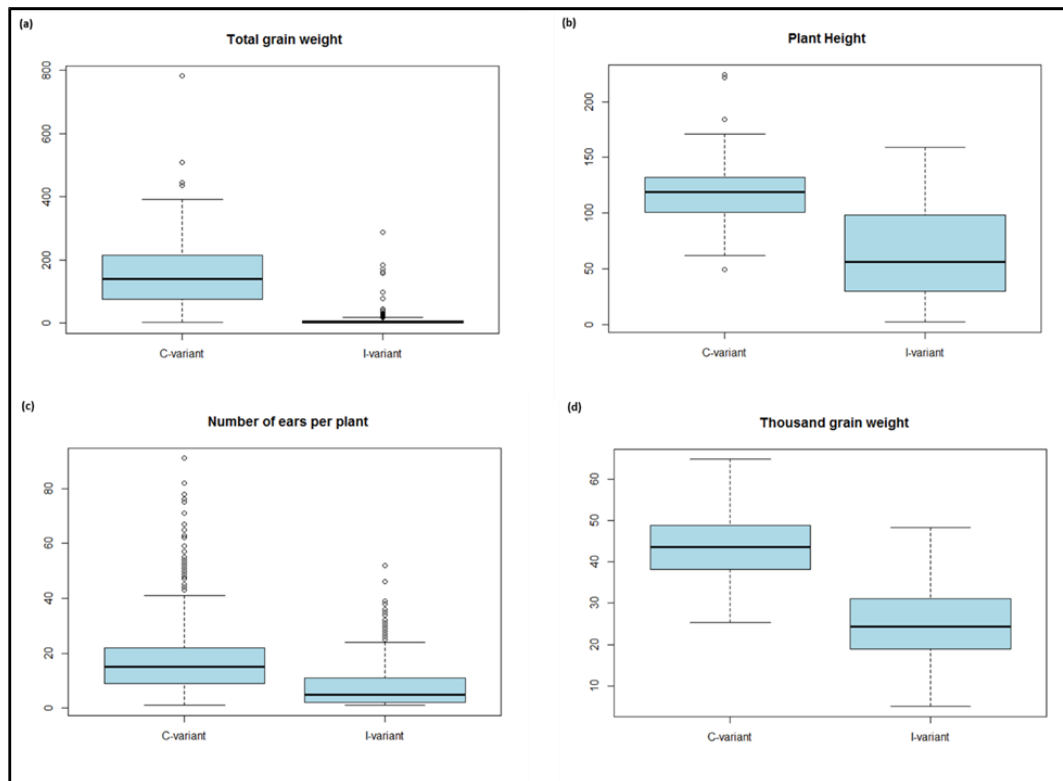[f] Coefficient of variation (in %)

Figure 1.   Boxplot based on genotype means for (a) total grain weight, (b) plant height, (c) number of ears per plant and (d) thousand grain weight.

The mean value was reduced by 92.8% under I-variant for ToGW. The lowest and highest values for ToGW was between 0 and 287.4 under I-variant and 2.2 and 781.7 under C-variant, respectively. The standard deviation (SD) was higher under C-variant (102.6) compared to I-variant (31.7). While, coefficient of variance (CV) showed the lower value under C-variant (65.7%) relative to I-variant (280.7%). The mean value was decreased 46.4% for HEI under I-variant. The 1 and 49 as minimum and 159 and 224 as maximum values could be observed for HEI under I-variant and C-variant, respectively. In addition, lower value for SD (20.3) and CV (17.2%) was observed for this trait under C-variant compared to I-variant. The mean value of NEP was reduced by 62.6 % under I-variant. The minimum value was zero and 1 under I-variant and C-variant, respectively. While, maximum value for this trait under I-variant and C-variant was 52 and 117, respectively. The value of SD was higher for NEP under C-variant (12.1) relative to I-variant (8.2). However, CV was lower under C-variant (65.8%) for this trait. The higher mean value (58.6%) was observed for TGW under C-variant relative to I-variant. The minimum value by 5 and 25.37 was observed for TGW under I-variant and C-variant, respectively. 48.3 g and 64.8 g were found as maximum values for this trait under I-variant and C-variant, respectively. The lower value of SD (7.8) and CV (17.9%) was estimated in the C-variant for TGW. The Shapiro-Wilk result indicates the phenotypic data of all four traits significantly deviated from the normal distribution.

Significant genotype effects (P<0.001) with degrees of freedom (DF) = 190 was observed for HEI and TGW using model 1. In addition, significant (P<0.001) treatment effects between I-variant and C-variant for ToGW, HEI, NEP and TGW was observed with applying model 1. Genotype-by-treatment interaction was significant (P<0.001) for HEI (DF =186), NEP (DF =186) and TGW (DF =184, Table 3). The significant genotype effect was observed for ELISA-60 using model 2.

Table 3. ANOVA result of all five measured traits for 191 investigated wheat genotypes.

Trait	Effect	Degrees of freedom	F Value	Pr > F
ELISA-60	Genotype (G)	190	2.26	<.0001
Total grain weight	Genotype (G)	190	1.05	0.49
	Treatment (T)	1	364.08	<.0001
	G*T	184	0.84	0.74
Plant height	Genotype (G)	190	5.43	<.0001
	Treatment (T)	1	3651.82	<.0001
	G*T	186	9.26	<.0001
Number of ears per plant	Genotype (G)	190	1.04	0.50
	Treatment (T)	1	443.86	<.0001
	G*T	186	1.83	<.0001
Thousand grain weight	Genotype (G)	190	3.05	<.0001
	Treatment (T)	1	1233.26	<.0001
	G*T	184	2.19	<.0001

3.2. Genotypic data

Genotyping of the 191 genotypes resulted in a raw marker set of 44,040 SNP. In total, 9,632 markers were excluded from the marker set, because of missing values, MAF < 5%, a heterozygosity >12.5%, or because markers were not uniformly distributed across chromosomes. The number of markers per chromosome was between 3,886 and 6,335. The minimum and maximum number of markers were found on chromosome 4H and 5H, respectively (Figure 2). Based on LD prune, a set of 3,117 markers was selected equally distributed on all seven barley chromosomes. The set of informative markers was used to estimate Kinship and population structure for identifying significant marker and QTL using GWAS.
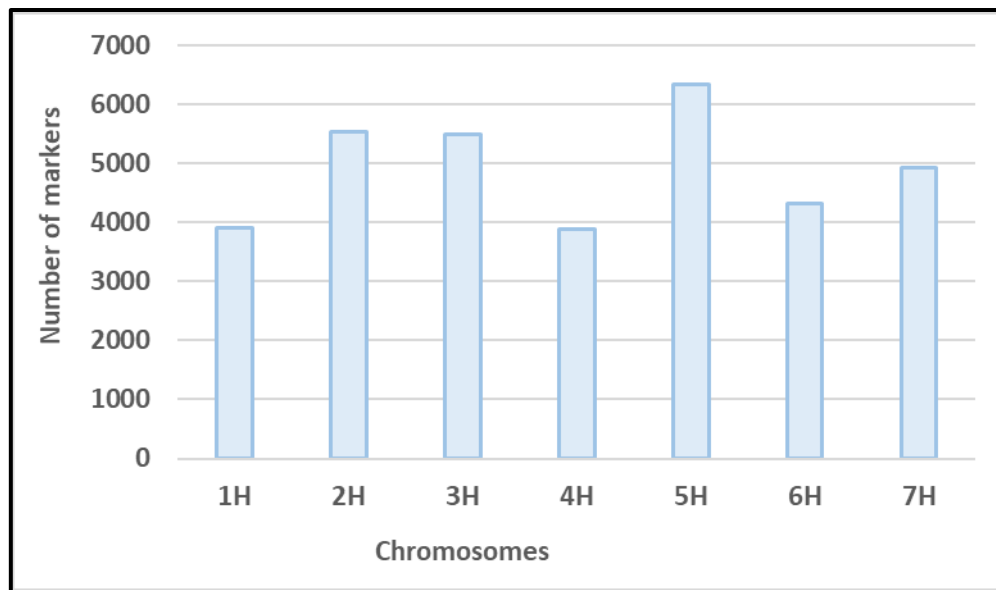
Figure 2. Distribution of SNP markers across all seven barley chromosomes.

3.2.1. Population structure

The conducted Bayesian cluster analysis revealed a number of K=2 subpopulations (Figure S1). Genotypes with a membership coefficient ≥0.7 to one of the Structure groups, were assigned to the corresponding group. Genotypes with a membership coefficient <0.7 were considered as admixed. Ninety, 77 and 24 genotypes were assigned to structure group 1 or 2 or the admixed group, respectively. Additionally, to visualize the results of the Bayesian cluster analysis, the structure grouping was projected on the results of PCoA (Figure 3). PCo1 and PCo2 explain 11.0% and 5.7% of the whole variation. In addition, genotypes were assigned to groups based on origin and row type (two and six rows) to define a clear connection between genotypes within a cluster. However, we could not identify a relation between genotypes based on the mentioned parameters within clusters. For instance, genotypes with two rows were distributed across K1, K2 and the admixture cluster. Furthermore, distribution of resistant and tolerant genotypes is shown in Figure 3. Resistant genotypes were clustered in K2. While, majority of tolerant genotypes (53%) were clustered in K1 and remaining genotypes (41% and 6%) were clustered in K2 and admixture clusters, respectively (Figure 3).
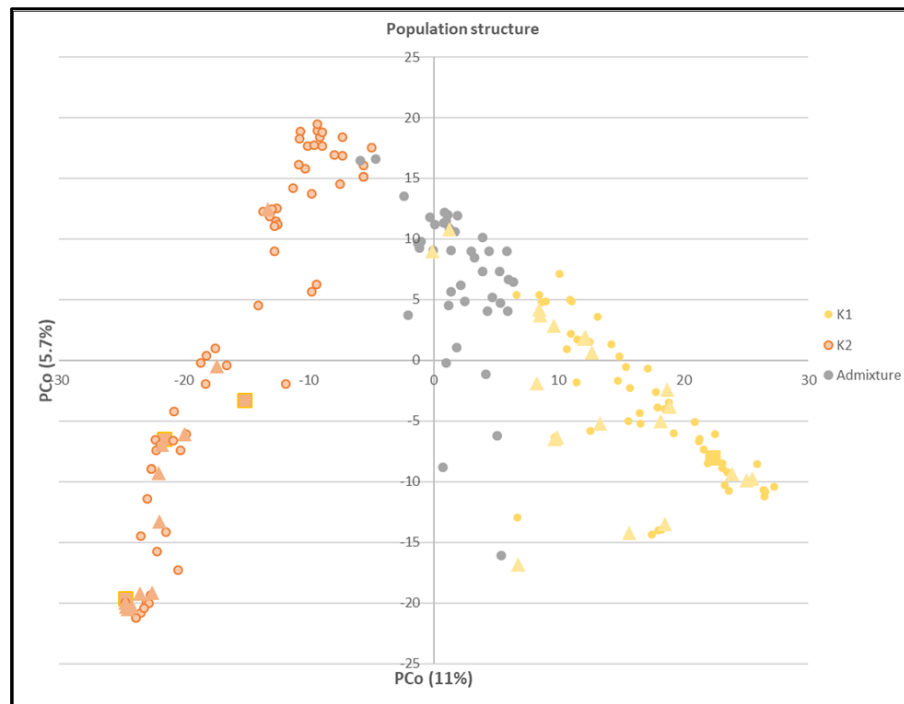
Figure 3. Principal coordinate analysis (PCoA) according to Structure grouping of 191 barley genotypes. Legend: yellow dots: genotypes assigned to Structure group 1, orange dots: genotypes assigned to Structure group 2 and gray dots: genotypes assigned to the admixed group. The rectangle and triangle indicates resistant and tolerant genotypes, respectively.

3.2.2. Genome-wide association study (GWAS):

To identify markers significantly associated with WDV, three different programs, i.e., TASEEL, GAPIT and FARMCPU, were used to conduct GWAS with three different models. A LOD value ≥3 was considered as threshold to identify significant marker-trait associations. Only markers that were identified by all three models, were considered as significant associations for the respective traits. In total, a number of nine significant associated markers with LOD ≥3 (-log10 of P value) were identified, which are partly distributed differently among the traits ELISA-60 (1), relative performance of ToGW (3), HEI (2), NEP (2) and TGW (1), respectively (Table 4 and Figure 4) based on relative trait values.

Table 4. List of common identified significant associated trait marker through three different used methods.

Trait	Marker name	Chr[a]	Pos[b]	P value			Identified genes in QTL region
				Gapit[c]	Tas-sel[c]	Farm-CPU	
ELISA-60	**JHI-Hv50k-2016-202912**	**3H**	**562758917**	3.3E-04	2.6E-04	2.2E-04	
Relative total grain weight	JHI-Hv50k-2016-196649	3H	534052013	6.8E-05	7.7E-04	2.8E-06	
Relative total grain weight	BOPA1_2955-452	4H	552300974	9.5E-04	2.4E-05	9.4E-05	*Cysteine proteinase inhibitor*
Relative total grain weight	BOPA2_12_10333	5H	554416618	3.4E-04	1.1E-04	1.7E-04	
Relative plant height	BOPA2_12_21049	2H	31329721	3.3E-05	3.5E-05	6.6E-04	*Dihydrofolate reductase*
Relative plant height	JHI-Hv50k-2016-435708	7H	1402273	1.5E-05	7.4E-05	1.1E-06	
Relative number of ears per plant	JHI-Hv50k-2016-123144	2H	631278948	1.8E-04	8E-04	2.7E-06	*NBS-LRR disease resistance protein*
Relative number of ears per plant	JHI-Hv50k-2016-142550	2H	666139797	6.4E-05	1.0E-04	1.7E-07	*Dihydroflavonol 4-reductase*
Relative thousand grain weight	JHI-Hv50k-2016-435708	7H	1402273	4.8E-06	1.7E-05	8.7E-09	

[a] Chromosome
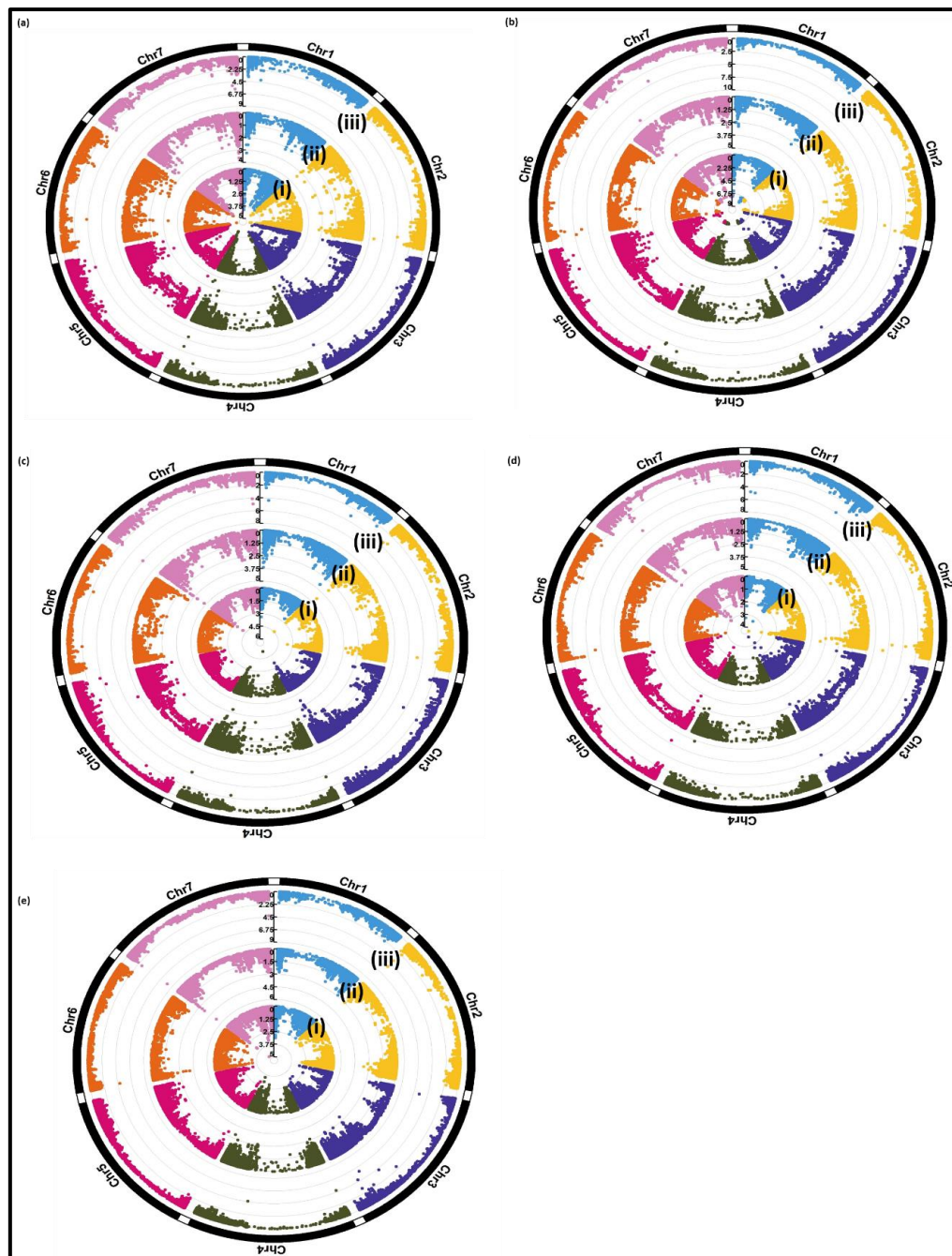[b] Position
[c] CMLM was applied in GAPIT and TASSEL

Figure 4. Circular Manhattan plot indicating three different GWAS methods: (i) CMLM (in TASSEL), (ii) CMLM (in GAPIT) and (iii) FarmCPU for (a) ELISA-60, (b) relative total grain weight (c) relative plant height, (d) relative number of ears per plant, (e) relative thousand grain weight.

An increased number of markers for ELISA-60 were identified by using the three programs TASSEL (596), GAPIT (26) and FARMCPU (41), located on all barley chromosomes. Chromosome 2H and 4H indicated the highest and lowest number of significant markers for ELISA-60. The phenotypic variance explained varied between 6.9% (JHI-Hv50k-2016-92202; on chromosome 2H) and 12.2% (JHI-Hv50k-2016-377967; on chromosome 6H). GAPIT identified 26 significantly associated markers for the trait ELISA-60 on all barley chromosomes. Chromosome 2H indicated the highest number of significantly associated markers (11 markers), of which three were also found by TASSEL. The 41 identified markers, identified by FARMCPU, were located on all barley chromosomes with five and three overlapping common markers with TASSEL and GAPIT, respectively. The marker "JHI-Hv50k-2016-202912" at a physical position of 562758917 bp on chromosome 3H was identified as common

marker among all three methods. Common markers that were identified by two or three different methods are shown in Figure S2 and Table S3.

For relative total grain number, 1184, 17 and 17 markers were identified on chromosomes 1H, 2H, 3H, 4H, 5H and 7H with TASSEL, GAPIT and FARMCPU, respectively. The set of 1184 identified markers with TASSEL were located on all barley chromosomes. Chromosome 7H and 4H indicated the highest and lowest number of significantly identified markers with 254 and 127 markers for relative total grain number, respectively. The phenotypic variation explained varied between 6.6% (JHI-Hv50k-2016-338274; on chromosome 5H) and 24% (JHI-Hv50k-2016-486135; on chromosome 7H). GAPIT identified 17 markers for relative total grain number on chromosome 2H, 3H, 4H, 5H and 7H. Five common markers were identified by GAPIT and TASSEL on chromosomes 3H, 4H and 5H. The same number of common markers was identified by GAPIT and FARMCPU. Three markers "JHI-Hv50k-2016-196649", "BOPA1_2955-452" and "BOPA2_12_10333" on chromosome 3H (at a physical position of 534052013 bp), 4H (at a physical position of 552300974 bp) and 5H (at a physical position of 554416618 bp) were detected by all three methods. In addition, five significant markers were identified by two methods.

For relative plant height, 14, 46 and 14 markers on chromosome 1H, 2H, 3, 4H, 5H and 7H were identified by TASSEL, GAPIT and FARMCPU, respectively. GAPIT identified a high number of significant markers for relative plant height compared to TASSEL and FARMCPU. The markers identified by GAPIT were distributed on all seven barley chromosomes. The chromosome 2H and 5H revealed the highest and lowest number of markers (10 and 1 marker(s)) for this trait. GAPIT showed 11 common markers with TASSEL on chromosome 1H, 2H, 3H, 4H and 7H. While, three out of 46 identified markers were common between GAPIT and FARMCPU on chromosome 1H, 2H, 5H and 7H. TASSEL identified 14 significant markers on chromosome 1H, 2H, 3H, 4H and 7H, which explaining phenotypic variance ranging from 7.44% (on chromosome 2H) to 14.18% (on chromosome 4H). FARMCPU identified 14 significant markers on all barley chromosomes with one exception (chromosome 6H). FARMCPU indicated two common markers with TASSEL on chromosome 2H and 7H. In total, twelve markers were identified by two methods and two out of these, "BOPA2_12_21049"and "JHI-Hv50k-2016-435708", were in addition identified by all three methods. These markers are located at a physical position of 31329721 bp and 1402273 bp on chromosome 2H and 7H, respectively. The marker "JHI-Hv50k-2016-435708" revealed the highest LOD value compared to other identified markers (Table S3).

In total, 8, 12 and 26 markers were significantly associated to the relative number of ears per plant using TASSEL, GAPIT, FARMCPU, respectively. TASSEL detected eight markers on four barley chromosomes (1H, 2H, 3H and 7H) which explained a phenotypic variance of 6.9% and 10.9% on chromosomes 7H (JHI-Hv50k-2016-439186) and 3H (JHI-Hv50k-2016-224192), respectively. GAPIT identified 12 markers significantly associated to the relative number of ears per plant which three (on chromosome 2H and 7H) and four (on chromosome 2H, 5H and 7H) common markers detected by TASSEL and FARMCPU, respectively. A set of 26 significant markers was identified using FARMCPU for relative number of ears per plant out of which three   were also identified by TASSEL on chromosome 1H and 2H.   Two markers "JHI-Hv50k-2016-123144" and "JHI-Hv50k-2016-142550" at a physical position of 631278948   and 666139797 on chromosome 2H were identified as common markers by all three methods used.

Nineteen, 18 and 12 significant markers were identified using MLM (in TASSEL), CMLM (in GAPIT) and FARMCPU for relative thousand-grain weight, respectively. The 19 significant markers for relative thousand-grain weight were distributed on all barley chromosome with the exception of chromosome 2H. The marker "SCRI_RS_174419 " and "JHI-Hv50k-2016-435708 "revealed the highest (12.7%) and lowest (6.1%) phenotypic variation on chromosome 1H and 2H, respectively.   Eight common markers were detected by TASSEL and GAPIT on chromosome 3H, 4H, 6H and 7H. Only one common marker was identified by TASSEL and FARMCPU on chromosome 7H. GAPIT and FARMCPU showed three common markers for relative thousand-grain weight on chromosome 3H and 7H. One significantly associated marker (JHI-Hv50k-2016-435708) on chromosome 7H was detected by all used methods. The marker "JHI-Hv50k-2016-435708" was significantly associated with

relative plant height and relative thousand grain weight at a physical position of 1402273 bp on chromosome 7H.

Finally, all identified significantly associated markers (Table 4) were screened to identify potential genes of interest, which were located within a distance of ±1 million base pairs of the identified significant marker based upon the calculated LD decay across all chromosomes. Three genes, a Dihydrofolate reductase, a NBS-LRR disease resistance protein and a Dihydroflavonol 4-reductase, were identified at chromosome 2H. Furthermore, a gene coding for a "Cysteine proteinase inhibitor" was identified at a distance of 374 bp from "BOPA1_2955-452" on chromosome 4H.

4. Discussion

The rising temperature, e.g., in many parts of Europe, led to environmental conditions that promote the spread of pests such as the leafhopper species *Psammotettix alienus*, which acts as a vector for *Wheat dwarf virus* (WDV). WDV is a generalist cereal pathogen and to date no resistance resources have been described for barley, except the cultivar "Post" [2]. Phenotyping of genotypes to identify resistant/tolerant genotypes that is based upon work including insects and viruses is labor intensive, time consuming and subject to environmental fluctuations in case it involves field tests. Hence, the availability of molecular markers would enable rapid and reliable discrimination between resistant/tolerant and susceptible genotypes [37]. Only little knowledge of genetic factors controlling WDV and resistance source s in barley is present and only cv. "Post" was identified as resistant [2]. In contrast, information about genetic markers associated to WDV for wheat was reported recently by Buerstmayr and Buerstmayr [38] and Pfrieme et al. [21].

As described by Nygren et al. [3], WDV causes symptoms such as dwarfing, tufting, streaks of leaf chlorosis as well as reduced spike numbers. Together with the relative virus titre, these traits were used for phenotyping in the present study. We identified 32 genotypes that show tolerance and resistance respectively to WDV. With regard to the expression of resistance [39] we identified genotypes with quantitative resistance that reduces or delays disease development and genotypes with qualitative resistance, preventing plant infection. Three out of these genotypes ("Res1", "Res2"and "Res3") did not show any virus titre accompanied by the absence of virus symptoms indicating a qualitative resistance. These genotypes originated from Afghanistan and Iran and are considered as favorable source for improving resistance to WDV in barley.

Considering the problems of phenotyping such as the lack of repeated tests in different years due to the challenging phenotyping method, different GWAS models were used in parallel, single and multiple locus models, to increase the probability of finding a true effect and to confirm detected markers in order to achieve reliable marker trait associations. TASSEL and GAPIT are MLM based models, which are considered as single locus models. These models contain a one-dimensional genome scan which tests one marker at a time, iteratively for each marker in a data set. These methods cannot match the real genetic model of complex traits which are controlled by multiple loci simultaneously [40]. To overcome this problem and reduce false positives associations that are caused by kinship and population structure from single locus models, the multilocus association mapping models are recommended [40]. FARMCPU as multilocus model eliminates confounding factors by testing associated markers as covariates through Fixed Effect Model (FEM) and optimization on the associated covariate markers using Random Effect Model (REM) [14]. Furthermore, FARMCPU reduces false positive associations by using both fixed and random effect models [14]. In the present study, MLM and CMLM as single locus model and FARMCPU as multi locus model were used to identify significant associated markers and QTLs. Among these tested models, GAPIT performed better than FARMCPU and TASSEL by considering obtained QQ_plot based on P values (Figure S3).

Nine common markers for all three methods for five measured traits were identified in the present study. No common markers for the three methods were identified on chromosome 1H and 6H. In the present study, the marker "JHI-Hv50k-2016-435708" was associated with relative plant height and relative thousand grain weight on chromosome 7H. These two traits are controlled by several

genes and are positively correlated [41]. He, and et al. [41] reported eight markers on barley chromo-somes 2H and 5H that are associated with plant height and thousand grain weight. Although our marker is not located on the same physical position of identified SNPs in the previous study, it could indicate new SNPs with a pleiotropic effect on chromosome 7H.

The identified common markers (among all three methods, Table 4) were screened for candidate genes according to published functional gene annotations of Morex V2 [24], leading to the identifica-tion of three high confidence genes on chromosome 2H (BOPA2_12_21049, JHI-Hv50k-2016-123144 and JHI-Hv50k-2016-142550) and one high confidence on chromosome 4H (BOPA1_2955-452), re-spectively. As a potential candidate resistance gene, the *Dihydrofolate reductase (DHFR)* gene on chro-mosome 2H was found to be co-localized with "BOPA2_12_21049" marker, which was associated to relative plant height. This gene plays several important roles in cell metabolism, catalyzes the con-version of dihydrofolate to tetrahydrofolates (synthesis of 5,6,7,8-tetrahydrofolate) [42, 43] and may lead to tolerance based upon compensation of virus induced metabolic changes in its host. A second high confidence gene "*NBS-LRR disease resistance*" was identified on chromosome 2H and was co-located with the identified marker "JHI-Hv50k-2016-123144" that is associated to the relative number of ears per plant. This gene belongs to a large group of disease resistance genes (*R* genes), which are involved exclusively in a non-membrane bound form in qualitative resistance to different viruses in various host plants [44]. In addition to the two identified candidate genes on chromosome 2H, *Dihy-droflavonol 4-reductase* was identified as third gene at 491 bp distance from marker "JHI-Hv50k-2016-142550". This gene plays a role in flavonoid metabolism. It is involved in the production of anthocy-anins and proanthocyanidins [45]. Flavonoids were shown to have antiviral activity [46]. The identi-fied marker on chromosome 4H corresponds to a gene coding for a Cysteine proteinase inhibitor that is located in a distance of 374 bp from "BOPA1_2955-452" marker. Cysteine proteinase inhibitors were reported to increase plant resistance against pathogens and insects [47-50]. The increase of re-sistance to potyviruses by using cysteine proteinase inhibitors in transgenic tobacco plants was re-ported by Gutierrez-Campos, Torres-Acosta [48]. Furthermore, Carrillo et al. [50] indicated that the barley cysteine-proteinase inhibitor reduces the performance of two aphid species in artificial diets and transgenic Arabidopsis thaliana plants.

The identification of WDV resistant or tolerant genotypes as well as the understanding of the genetic background of plants is the prerequisite to reduce negative effects of this virus on plant pro-duction. In this context, identified markers or QTLs do not only provide a relevant genetic basis for breeding but also enhance our knowledge about genomic regions, which are controlling WDV re-sistance in barley.

5. Conclusions

We report about a first GWAS study that used the barley 50K iSelect chip to identify associated markers for resistance against WDV. In the present study, three different statistical models (MLM, CMLM, and FARM CPU) were used to identify significant marker trait association and QTL. While three genotypes show a qualitative resistance to WDV based upon phenotypic data, the GWAS anal-ysis of the total plant set of 191 genotypes allowed the identification of nine significantly associated markers. The development of KASP (Kompetitive allele-specific PCR) markers based on the obtained common significant markers could be a valuable tool for plant breeding to replace the classical screen-ing method including vector insects and virus.

Author Contributions

Conceptualization, Antje Habekuß and Frank Ordon; Data curation, Antje Habekuß and Torsten Will; Formal analysis, Behnaz Soleimani and Heike Lehnert; Funding acquisition, Antje Habekuß and Frank Ordon; Investigation, Behnaz Soleimani, Heike Lehnert, Sarah Trebing, Andreas Stahl and Torsten Will; Supervision, Torsten Will; Validation, Behnaz Soleimani, Heike Lehnert, Frank Ordon, Andreas Stahl and Torsten Will; Visualization, Behnaz Soleimani and Torsten Will; Writing – original draft, Behnaz Soleimani and Torsten Will; Writing – review & editing, Behnaz Soleimani, Heike Lehnert, Antje Habekuß, Frank Ordon, Andreas Stahl and Torsten Will.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Mayer, K.F.X., et al., *A physical, genetic and functional sequence assembly of the barley genome.* Nature, 2012. **491**(7426): p. 711-+.
2. Habekuß, A., et al., *Breeding for resistance to insect-transmitted viruses in barley-an emerging challenge due to global warming.* Journal für Kulturpflanzen, 2009. **6(12)**: p. 53-61.
3. Nygren, J., et al., *Variation in Susceptibility to Wheat dwarf virus among Wild and Domesticated Wheat.* Plos One, 2015. **10**(4).
4. Cejnar, P., et al., *Two mutations in the truncated Rep gene RBR domain delayed the Wheat dwarf virus infection in transgenic barley plants.* Journal of Integrative Agriculture, 2018. **17**(11): p. 2492-2500.
5. Vacke, J., *Host plants range and symptoms of wheat dwarf virus.* ĕVĕd Pr Výz Ústavú Rostl Výroby Praha-Ruzyn, 1972. **17**: p. 151-162.
6. Vacke, J. and R. Cibulka, *Reactions of registered winter barley varieties to wheat dwarf virus infection.* Czech Journal of Genetics and Plant Breeding-UZPI (Czech Republic), 2001.
7. Vacke, J., *Wheat dwarf virus.* Biol. Plant. (Praha), 1961. **3**: p. 228-233.
8. Koklu, G., J.N.E. Ramsell, and A. Kvarnheden, *The complete genome sequence for a Turkish isolate of Wheat dwarf virus (WDV) from barley confirms the presence of two distinct WDV strains.* Virus Genes, 2007. **34**(3): p. 359-366.
9. Kanzi, A.M., et al., *Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance.* Frontiers in Genetics, 2020. **11**.
10. Wang, S.C., et al., *Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array.* Plant Biotechnology Journal, 2014. **12**(6): p. 787-796.
11. Bradbury, P.J., et al., *TASSEL: software for association mapping of complex traits in diverse samples.* Bioinformatics, 2007. **23**(19): p. 2633-2635.
12. Purcell, S., et al., *PLINK: A tool set for whole-genome association and population-based linkage analyses.* American Journal of Human Genetics, 2007. **81**(3): p. 559-575.
13. Lipka, A.E., et al., *GAPIT: genome association and prediction integrated tool.* Bioinformatics, 2012. **28**(18): p. 2397-2399.
14. Liu, X., et al., *Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies* Plos One, 2016. **12**(3).
15. Tsai, H.Y., et al., *Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat (vol 10, 3347, 2020).* Scientific Reports, 2020. **10**(1).
16. Gyawali, S., et al., *Genome wide association studies (GWAS) of spot blotch resistance at the seedling and the adult plant stages in a collection of spring barley.* Molecular Breeding, 2018. **38**(5).

17. Novakazi, F., et al., *Genome-wide association studies in a barley (Hordeum vulgare) diversity set reveal a limited number of loci for resistance to spot blotch (Bipolaris sorokiniana).* Plant Breeding, 2020. **139**(3): p. 521-535.

18. Wehner, G.G., et al., *Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley.* Bmc Plant Biology, 2015. **15**.

19. Jabbari, M., et al., *GWAS analysis in spring barley (Hordeum vulgare L.) for morphological traits exposed to drought.* Plos One, 2018. **13**(9).

20. Rode, J., et al., *Identification of marker-trait associations in the German winter barley breeding gene pool (Hordeum vulgare L.).* Molecular Breeding, 2012. **30**(2): p. 831-843.

21. Pfrieme, A.K., et al., *Identification and validation of Quantitative Trait Loci for Wheat dwarf virus resistance in wheat (Triticum spp.) (vol 13, 828639, 2022).* Frontiers in Plant Science, 2022. **13**.

22. Bayer, M.M., et al., *Development and Evaluation of a Barley 50k iSelect SNP Array.* Frontiers in Plant Science, 2017. **8**.

23. Doyle, J.F. and J.L. Doyle, *Isolation of plant DNA from fresh tissue.* Focus 12, 1990. **12**: p. 13-15.

24. Mascher, M., *Pseudomolecules and annotation of the second version of the reference genome sequence assembly of barley cv. Morex [Morex V2].* 2019.

25. Browning, B.L. and S.R. Browning, *A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals.* American Journal of Human Genetics, 2009. **84**(2): p. 210-223.

26. Browning, S.R. and B.L. Browning, *Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.* American Journal of Human Genetics, 2007. **81**(5): p. 1084-1097.

27. Reif, J.C., A.E. Melchinger, and M. Frisch, *Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management.* Crop Science, 2005. **45**(1): p. 1-7.

28. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-959.

29. Perrier, X. and J.P. Jacquemoud-Collet, *DARwin software: Dissimilarity analysis and representation for windows. Website* http://darwin. *cirad. fr/darwin [accessed 1 March 2013].* 2006.

30. Earl, D.A. and B.M. Vonholdt, *STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.* Conservation Genetics Resources, 2012. **4**(2): p. 359-361.

31. R Core Team, *R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available online at:* http://www.R-project.org/. 2014.

32. Shin, J.H., et al., *LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms.* Journal of statistical software, 2006. **16**: p. 1-9.

33. Warnes, G., et al., *Genetics: Population Genetics. R package version 1.3.8.1. Available online at:* http://CRAN.R-project.org/package=genetics. . 2013.

34. Voss-Fels, K., et al., *Subgenomic Diversity Patterns Caused by Directional Selection in Bread Wheat Gene Pools.* Plant Genome, 2015. **8**(2).

35. Sannemann, W., et al., *Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept.* Molecular Breeding, 2015. **35**(3).

36. Lehnert, H., et al., *Genome-Wide Association Studies Reveal Genomic Regions Associated With the Response of Wheat (Triticum aestivum L.) to Mycorrhizae Under Drought Stress Conditions.* Frontiers in Plant Science, 2018. **9**.

37. Hasan, N., et al., *Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes.* J Genet Eng Biotechnol, 2021. **19**(1): p. 128.

38. Buerstmayr, M. and H. Buerstmayr, *Two major quantitative trait loci control wheat dwarf virus resistance in four related winter wheat populations.* Theoretical and Applied Genetics, 2023. **136**(5): p. 103.

39. Rousseau, E., et al., *Virus epidemics, plant-controlled population bottlenecks and the durability of plant resistance.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2019. **374**(1775).

40. Kaler, A.S., et al., *Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize.* Frontiers in plant science, 2020: p. 1794.

41. He, T., T.T. Angessa, and C. Li, *Pleiotropy structures plant height and seed weight scaling in barley despite long history of domestication and breeding selection.* Plant Phenomics, 2020.

42. Rao, K.N. and S.R. Venkatachalam, *Inhibition of dihydrofolate reductase and cell growth activity by the phenanthroindolizidine alkaloids pergularinine and tylophorinidine: the in vitro cytotoxicity of these plant alkaloids and their potential as antimicrobial and anticancer agents.* Toxicology in Vitro, 2000. **14**(1): p. 53-59.

43. Gorelova, V., et al., *Dihydrofolate Reductase/Thymidylate Synthase Fine-Tunes the Folate Status and Controls Redox Homeostasis in Plants.* Plant Cell, 2017. **29**(11): p. 2831-2853.

44.   Maule, A.J., C. Caranta, and M.I. Boulton, *Sources of natural resistance to plant viruses: status and prospects.* Molecular Plant Pathology, 2007. **8**(2): p. 223-231.

45.   Kristiansen, K.N. and W. Rohde, *Structure of the Hordeum vulgare gene encoding dihydroflavonol-4-reductase and molecular analysis of ant 18 mutants blocked in flavonoid synthesis.* Molecular and General Genetics MGG, 1991. **230(1)**: p. 49-59.

46.   Badshah, S.L., et al., *Antiviral activities of flavonoids.* Biomedicine & Pharmacotherapy, 2021. **140**.

47.   Masoud, S.A., et al., *Expression of a Cysteine Proteinase-Inhibitor (Oryzacystatin-I) in Transgenic Tobacco Plants.* Plant Molecular Biology, 1993. **21**(4): p. 655-663.

48.   Gutierrez-Campos, R., et al., *The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants.* Nature Biotechnology, 1999. **17**(12): p. 1223-1226.

49.   Bouchard, E., D. Michaud, and C. Cloutier, *Molecular interactions between an insect predator and its herbivore prey on transgenic potato expressing a cysteine proteinase inhibitor from rice.* Molecular Ecology, 2003. **12**(9): p. 2429-2437.

50.   Carrillo, L., et al., *A barley cysteine-proteinase inhibitor reduces the performance of two aphid species in artificial diets and transgenic Arabidopsis plants.* Transgenic Research, 2011. **20**(2): p. 305-319.