



OPEN ACCESS

EDITED BY

Jinyoung Y. Barnaby,
United States Department of
Agriculture (USDA), United States

REVIEWED BY

Sivakumar Sukumaran,
The University of Queensland,
Australia
Shuwei Liu,
Shandong University (Qingdao), China

*CORRESPONDENCE

Behailu Mulugeta
✉ behailu.mulugeta@slu.se

SPECIALTY SECTION

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

RECEIVED 01 August 2022

ACCEPTED 20 December 2022

PUBLISHED 26 January 2023

CITATION

Mulugeta B, Tesfaye K, Ortiz R,
Johansson E, Hailesilassie T,
Hammenhag C, Hailu F and Geleta M
(2023) Marker-trait association
analyses revealed major novel QTLs
for grain yield and related traits in
durum wheat.
Front. Plant Sci. 13:1009244.
doi: 10.3389/fpls.2022.1009244

COPYRIGHT

© 2023 Mulugeta, Tesfaye, Ortiz,
Johansson, Hailesilassie, Hammenhag,
Hailu and Geleta. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Marker-trait association analyses revealed major novel QTLs for grain yield and related traits in durum wheat

Behailu Mulugeta^{1,2,3*}, Kassahun Tesfaye^{1,4}, Rodomiro Ortiz²,
Eva Johansson², Teklehaimanot Hailesilassie¹,
Cecilia Hammenhag², Faris Hailu⁵ and Mulatu Geleta²

¹Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia, ²Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden, ³Sinana Agricultural Research Center, Oromia Agricultural Research Institute, Bale-Robe, Ethiopia, ⁴Director General, Bio and Emerging Technology Institute (BETin), Addis Ababa, Ethiopia, ⁵Department of Biology and Biotechnology, Wollo University, Dessie, Ethiopia

The growing global demand for wheat for food is rising due to the influence of population growth and climate change. The dissection of complex traits by employing a genome-wide association study (GWAS) allows the identification of DNA markers associated with complex traits to improve the productivity of crops. We used GWAS with 10,045 single nucleotide polymorphism (SNP) markers to search for genomic regions associated with grain yield and related traits based on diverse panels of Ethiopian durum wheat. In Ethiopia, multi-environment trials of the genotypes were carried out at five locations. The genotyping was conducted using the 25k Illumina Wheat SNP array to explore population structure, linkage disequilibrium (LD), and marker-trait associations (MTAs). For GWAS, the multi-locus Fixed and Random Model Circulating Probability Unification (FarmCPU) model was applied. Broad-sense heritability estimates were high, ranging from 0.63 (for grain yield) to 0.97 (for thousand-kernel weight). The population structure based on principal component analysis, and model-based cluster analysis revealed two genetically distinct clusters with limited admixtures. The LD among SNPs declined within the range of 2.02–10.04 Mbp with an average of 4.28 Mbp. The GWAS scan based on the mean performance of the genotypes across the environments identified 44 significant MTAs across the chromosomes. Twenty-six of these MTAs are novel, whereas the remaining 18 were previously reported and confirmed in this study. We also identified candidate genes for the novel loci potentially regulating the traits. Hence, this study highlights the significance of the Ethiopian durum wheat gene pool for improving durum wheat globally. Furthermore, a breeding strategy focusing on accumulating favorable alleles at these loci could improve durum wheat production in the East African highlands and elsewhere.

KEYWORDS

candidate gene, durum wheat, GWAS, linkage disequilibrium, population structure, QTL

1 Introduction

Durum wheat (*Triticum turgidum*, L. var. *durum* Desf.) is a staple cereal crop produced to make pasta, bread, and other traditional food items (Sall et al., 2019; Ceglar et al., 2021). Durum wheat accounts for approximately 8% of global wheat production (Sall et al., 2019), and most of it (75%) is produced in the Mediterranean region (Xynias et al., 2020). The world's largest producers are Turkey and Canada, while Ethiopia is the largest producer in Sub-Saharan Africa (SSA). Durum wheat was domesticated around 10,000 years ago in the Fertile Crescent (Özkan et al., 2002; Soriano et al., 2018; Ceglar et al., 2021) and has since then been a vital source of energy, minerals, and bioactive compounds in human nutrition (Johansson et al., 2020a). The durum wheat is an amphidiploid species containing an AABB genome, and its genome size is nearly 12 Gb (Maccaferri et al., 2019).

The current development of advanced DNA sequencing methods, functional genomic tools, and availability of different DNA chip technology has highly facilitated the genetic dissection of multi-genic traits of food crops (Collard and Mackill, 2008; Al-Khayri et al., 2016; Geleta and Ortiz, 2016). Association mapping (AM) has been widely used to dissect the genetic architecture behind traits like grain yield, host plant resistance to pathogens, drought and salinity tolerance, phenology, and quality traits (Maccaferri et al., 2010; Tuberosa, 2012; Canè et al., 2014; Turki et al., 2015; Giraldo et al., 2016; Mengistu et al., 2016; Kidane et al., 2019; Mérida-García et al., 2019; Mérida-García et al., 2020). Moreover, genome-wide association studies (GWAS) have been successfully used to map genetic loci and dissect the genomic regions underlying several vital traits in important food crops, such as barley (Bellucci et al., 2017; Borrego-Benjumea et al., 2021), and bread wheat (Li et al., 2019; Gao et al., 2021; Mekonnen et al., 2021).

In wheat, GWAS has been successfully applied to identify and dissect QTL associated with grain yield (Li et al., 2019; Gao et al., 2021), host plant resistance to pathogens (Alemu et al., 2021a; Alemu et al., 2021; Mekonnen et al., 2021), drought tolerance (Bhatta et al., 2018; Mathew et al., 2019), root architecture (Alemu et al., 2021b), phenology (Mekonnen et al., 2021), adaptation to salinity (Quamruzzaman et al., 2021), and end-use quality traits (Chen et al., 2019; Talini et al., 2020). However, in durum wheat, limited GWAS results have been reported across traits of interest, although some results are present for grain yield (Wang et al., 2019; Anuarbek et al., 2020), host plant resistance to pathogens (Liu et al., 2017b; Aoun et al., 2021), drought tolerance (Wang et al., 2019), root system architecture (Maccaferri et al., 2016; Alemu et al., 2021b), osmotic adjustment (Condorell et al., 2022), and phenology and quality traits (Fiedler et al., 2017). Furthermore, GWAS results reported on Ethiopian durum wheat cultivars and

landraces are insufficient. Increased genomic research is needed to improve durum wheat production in Ethiopia by utilizing genomic-assisted breeding approaches.

Wheat landraces can be seen as an essential germplasm resource, with the potential to be utilized as a reservoir of crop diversity that harbors significant novel loci associated with agronomic, phenological, and end-use quality traits (Johansson et al., 2021). Landraces and their wild relatives have served as sources of valuable genes to improve modern cultivars for adaptation to diverse environments, grain yield, end-use quality, host plant resistance to the pathogen, and abiotic stress tolerance (Maccaferri et al., 2019; Johansson et al., 2020b; Sansaloni et al., 2020). Several reports revealed that Ethiopian durum wheat has high genetic diversity to be explored in the search for essential novel and valuable genes for improvements of traits such as grain yield, nutritional quality, host plant resistance to pathogens, and drought tolerance (Mengistu et al., 2016; Kabbaj et al., 2017; Kidane et al., 2019; Alemu et al., 2020a). Hence, understanding the genetic basis of these important traits using recent genomic-based research will facilitate the use of Ethiopian germplasm in an improvement program to maintain a food-secure future in the region.

This study aimed to use GWAS to define genomic regions in Ethiopian durum wheat associated with grain yield and related traits. Furthermore, population structure and linkage disequilibrium were evaluated for precise identification of the genetic basis of valuable genomic regions associated with grain yield and important agronomic traits.

2 Materials and methods

2.1 Germplasm

The present study used 420 Ethiopian durum wheat landraces and cultivars. To accommodate the extensive diversity of the Ethiopian durum wheat gene pool, 385 landraces were selected from different geographical regions of Ethiopia, while 35 were crossbred cultivars. [Supplementary Table 1](#) provides information on these landraces and cultivars. For simplicity, the landraces and cultivars will be designated as genotypes hereafter.

2.2 Description of test environment

The performance of the genotypes was evaluated across five locations in Ethiopia, namely, Akaki (AK; 09°53' 48" N/39°49' 16" E), Chefe Donsa (CD; 08°58' 57" N/39°09' 13" E), Holeta (HO; 09°01' 15" N/38°28' 26" E), Kulumsa (KU; 08°01' 11" N/39°09' 37" E) and Sinana (SN; 07°06' 58" N/40°13' 38" E)

during the 2019–2020 main crop-growing season. The testing locations represent the country's major and most suitable durum wheat growing environments. The soil texture of each site is characterized as heavy clay for Akaki and Chefe Donsa and clay for Holeta, Kulumsa, and Sinana. The test sites are classified into ME (Mega environment)2:SW (Spring Wheat) high rainfall areas that receive more than 500 mm of rainfall during the crop growing cycle as defined by CIMMYT's Wheat Breeding Program (Rajaram et al., 1994). Among the five test sites, Sinana, Kulumsa, and Holeta have been used by CIMMYT's wheat breeding program targeting high potential environments in the highlands of East Africa. The Agro-ecology at the Akaki and Chefe Donsa sites are also similar to those at the other three test sites and are considered high-potential sites. During the crop-growing season, the mean monthly maximum and minimum temperature of the Sinana site ranged from 20.8°C to 23.9°C and 8.4°C to 9.2°C, respectively, with total rainfall of 810 mm (Supplementary Table 2). The Holeta site received a total rainfall of 852 mm, with the mean annual minimum and maximum temperature of 10°C and 24°C, respectively (Supplementary Table 2). The Chefe Donsa site received mean monthly minimum and maximum temperatures ranging from 9.4–11.8°C and 20.3–24.2°C, with a total rainfall of 870.5 mm. Whereas, the Akaki site received mean monthly minimum and maximum temperatures ranging from 9.89–13.82°C and 24.85–26.91°C, respectively, with a total rainfall amount of 711.5 mm. Kulumsa site received a total rainfall of 700 mm, with a mean monthly minimum temperature of 11°C and a mean monthly maximum temperature of 23°C.

2.3 Field experimental design

The experiment was laid out using an alpha lattice design with two replications containing 21 incomplete blocks with a block size of 20, according to Patterson and Williams (1976). The landraces and cultivars were randomly assigned and planted on a plot size of 1 m² with 2.5 m x 0.4 m (two rows with 20 cm spacing). The space between the plots was 20 cm. A seed rate of 150 kg ha⁻¹ and fertilizer rate of 50 kg N ha⁻¹ and 100 kg of P₂O₅ ha⁻¹ was applied to each plot. In order to maintain genotype uniformity (since the genotypes were mostly landraces with possible seed admixture), the genotypes were grown on different plots for two consecutive crop growing seasons (2017–2018) at Sinana agricultural research center, and individual plants that appeared to differ in any of the clearly visible phenotypic traits were removed.

2.4 Evaluation of phenotypic traits

In this study, phenotyping was conducted by applying the previously described methodology for evaluating wheat genetic

resources (IBPGR, 1985). The traits measured were phenology (days to heading, days to physiological maturity, and grain filling period), plant architecture (plant height, spike length, and number of effective tillers per plant), grain yield, and grain yield-related traits (number of spikelets per spike, and thousand kernel weight).

2.5 Statistical analysis of the phenotypic data

Before further analysis, data were evaluated by the Shapiro–Wilk test to assess if they fit into the normal distribution. Furthermore, based on the results from the normality test, the homogeneity test was performed for the scored data in the experiment as described in Levene (1960). The R statistical software (R Development Core team, 2021) was used for computing descriptive statistics (mean, range, standard deviation), coefficient of variation, analysis of variance (ANOVA), correlation among traits, and broad-sense heritability. The linear mixed model (LMM) fitted by the Restricted/Estimated Maximum Likelihood method [REML, Corbeil and Searle (1976)] in R package “lme4” (Bates et al., 2015) was used to estimate the variance components of scored traits. To perform ANOVA for each test environment, the genotypes and blocks were considered fixed and random effects, respectively. The response of the *i*th genotype in the *j*th incomplete block with the *l*th replication of each environment for a particular trait was described as:

$$Y_{ijl} = \mu + \tau_i + \beta_j + \gamma_{l(j)} + \xi_{ijl} \quad (1)$$

where Y_{ijl} is the phenotypic response of the *i*th genotype in *l*th incomplete block within *j*th replication, μ is the overall mean, τ_i is the fixed effect of genotype *i*, β_j is the random effect of the *j*th replicate, γ_l is the random effect of the *j*th incomplete block nested in the *l*th replication, and ξ_{ijl} is the residual error.

The combined ANOVA across environments inference was computed for all the response variables as follows:

$$Y_{ijkl} = \mu + \tau_i + \beta_l + \gamma_j + E_k + \beta\gamma_{lj} + \gamma E_{jk} + \tau E_{ik} + \xi_{ijkl} \quad (2)$$

where Y_{ijkl} is the observed phenotypic trait for *i*th genotype in *l*th incomplete block within *j*th replication at the *k*th environment, μ is the overall mean, τ_i is the fixed effect of genotype *i*, β_l is the random effect of *j*th replication, γ_j is the random effect of the *l*th incomplete block within *j*th replication, E_k is the random effect of environment *k*, $\beta\gamma_{lj}$ is random effect of incomplete block *l* nested in replication *j*, γE_{jk} is random effect of replication *j* in test environment *k*, τE_{ik} is random effect of interaction between genotype *i* and environment *k*, and ξ_{ijkl} is a random residual effect. For the sake of simplicity, we assumed that all the underlying random effects residuals are normally distributed with zero mean and are independent homoscedastic.

The best linear unbiased estimates (BLUEs) of measured traits for each genotype from each environment were obtained using META R software (Alvarado et al., 2020). The estimated means of BLUEs was used to compute the Pearson correlation coefficient (r) by the “cor. test” function in the R (R Development Core team, 2021) and GWAS analysis. The estimates of broad-sense heritability (H^2) were computed from pooled ANOVA across environments (Gonçalves-Vidigal et al., 2008) as:

$$H^2 = \frac{\sigma_g^2}{[\sigma_g^2 + (\sigma_{gl}^2/l) + (\sigma_e^2/lr)]} \quad (3)$$

where σ_g^2 is genotypic variance, σ_{gl}^2 is genotype by environment interaction variance, σ_e^2 is environmental variance, l is the number of environments, and r is the number of replications.

2.6 DNA extraction, genotyping, and filtering of SNP markers

A single spike representing each genotype was collected during field phenotyping for genotyping. Five healthy seeds from each spike were taken to represent each genotype and were planted in 3 L pots in a greenhouse at the Swedish University of Agricultural Science (SLU), Alnarp, Sweden. A total of ten 6 mm leaf discs sampled from five two-week-old seedlings of each genotype were collected in each well of a 96-deep well plate and freeze-dried using the CoolSafe ScanVAC Freeze Dryer according to the instructions provided by Trait Genetics. The freeze-dried samples in 96-well deep well plates were sent to TraitGenetics (GmbH, Gatersleben, Germany) for DNA extraction and subsequent genotyping. A standard cetyltrimethylammonium bromide (CTAB) protocol was used to extract DNA from the leaf samples in TraitGenetics' lab. The 420 genotypes were genotyped using an Illumina Infinium 25k wheat single nucleotide polymorphism (SNP) array following the manufacturer's protocol. The details of the SNP array can be found at <https://www.traitgenetics.com/index.php/service-products>. Based on a specific durum wheat cluster file developed by TraitGenetics that differentiates durum wheat from bread wheat, markers accurately scored for the A and B genomes were recorded.

Several criteria were used to filter the genotypic data obtained before further analysis. TASSEL 5.2.80 software (Bradbury et al., 2007) was used to remove SNP loci with missing data above 5% or with minor allele frequency (MAF) below 5% (including monomorphic loci). Further filtering of the remaining SNP loci was conducted based on the level of observed heterozygosity (H_o), and loci with H_o greater than 0.01 were excluded. These filtering steps resulted in 10,045 SNPs that were used for data analyses. The evaluation of these SNP loci showed that each of the 420 samples had less than 1%

missing data, and hence no genotype was excluded from the data analyses.

2.7 Population structure and linkage disequilibrium (LD) analysis

The number of subgroups among the 420 genotypes was inferred by principal component analysis (PCA) and model-based clustering methods, which were computed by Genome Association and Integrated Prediction Tool (GAPIT) 3.0 (Wang and Zhang, 2021) and STRUCTURE 2.3.4. software (Pritchard et al., 2000; Falush et al., 2007), respectively. A Bayesian approach (MCMC: Markov Chain Monte Carlo) that assumes an ancestry model of ADMIXTURE and correlated allele frequencies among the subgroups was used for model-based cluster analysis. The length of the burn-in period was adjusted to 50,000, followed by 100,000 MCMC iterations for subgroups (K) ranging from one to ten. Ten independent runs were carried out for each K . The STRUCTURE results were visualized using STRUCTURE Harvester (Earl and vonHoldt, 2012). The number of best K was inferred using the delta K method described in Evanno et al. (2005). The optimum K value bar plot was drawn based on CLUMPAK online software (Kopelman et al., 2015).

Information on the pattern of linkage disequilibrium (LD) within a genetic material of interest is necessary to determine the marker density required for a genome-wide scan (Siol et al., 2017). Accordingly, LD was computed using TASSEL version 5.2.8 (Bradbury et al., 2007). The pairwise LD (squared allele frequency, r^2) for pairs of SNP markers was computed according to Weir (1997). The intersection of the fitted curve with the cut-off threshold was considered the mean r^2 value for each chromosome (Bresghehlo and Sorrells, 2006b; Liu et al., 2017c). The mean r^2 value of each chromosome was computed and plotted against the chromosome's physical distance. The physical distance at which the r^2 value dropped to half its average maximum value was considered the LD decay rate (Huang et al., 2010). The $r^2 = 0.3$ ($p < 0.01$) was considered as a cut-off point to represent a limit of QTL between pairs of markers as indicated in previous studies for Ethiopian durum wheat panels (Liu et al., 2017c; Alemu et al., 2021b).

2.8 Identification of marker trait association

GWAS was conducted using best linear unbiased estimates (BLUEs) for nine phenotypic traits and 10,045 SNP markers to identify marker-trait association (MTA). The BLUEs for grain yield, spike length, and grain-filling period were calculated by considering days to heading (DTH) as a covariate in order to control the effect of heading time, as suggested in previous studies (Sabadin et al., 2012; Tuberosa, 2012). The analysis was

performed by employing a multi-locus-based method, fixed and random model Circulating Probability Unification [FarmCPU, Liu et al. (2016)] model selection algorithm implemented in GAPIT 3 R package (Lipka et al., 2012; Tang et al., 2016; Wang and Zhang, 2021). The FarmCPU model algorithm eliminates potential confounding factors by employing the fixed and random effect models iteratively. This was done to overcome the overfitting model influences of the stepwise regression and to control spurious MTA caused by population structure and family relatedness. GAPIT 3 was also used to visualize the Manhattan and Quantile-quantile (QQ)-plots. The QQ-plot fits the model to account for the population structure.

The stringent false-positive discovery rate [FDR, $p < 0.01$ (Benjamini and Hochberg, 1995)] and Bonferroni-corrected threshold of $(-\log_{10}(0.05/n)) = 5.30$ was used, where n is the total numbers of SNPs) to declare a significant MTA between a marker and phenotypic trait. All MTAs above the threshold levels were rated as significant. The percentage of phenotypic variance explained (PVE) by individual MTA (Garcia et al., 2019) and a marker-based VanRaden kinship (K) matrix (VanRaden, 2008) for the genotypes of interest was also generated in R/GAPIT 3. It was assumed that an identified QTL is stable in the genomic region when significant MTA has appeared in two or more test locations, and the additive effects were concordant.

2.9 Identification of putative novel MTAs and associated candidate genes

The novelty of significant MTAs and their potential associated genes were determined by comparative analyses with previously published reports using different *Triticum* databases such as GrainGene, T3/wheat, and Wheat URGI (Alaux et al., 2018). The lists of different genes and functions were downloaded from the NCBI database (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/231/445/GCA_900231445.1_Svevo.v1/) to identify genes related to significant MTAs. The nucleotide position extending from 1-5 cM up and downstream from the SNP position was used for searching the potential candidate genes, as previously reported for wheat (Breseghello and Sorrells, 2006a). The genes associated with the significant MTAs were obtained from the durum wheat (*Triticum turgidum* (Svevo.v1) reference genome) (Maccaferri et al., 2019).

3 Results

3.1 Phenotypic mean performance of genotypes

Descriptive statistics, frequency distribution, and boxplots clearly showed a wide range of variation for all the traits evaluated (Table 1 and Figure 1). The mean number of days for

days to heading, days to physiological maturity, and grain filling period over the combined environments were 72.6, 136, and 63.4, respectively (Table 1). The mean grain yield was 6.7 t ha^{-1} , while the mean thousand kernel weight was 40.9 g. On average, 17.5 spikelets per spike were recorded across the environments (Table 1). The highest mean grain yield (8.4 t ha^{-1}) was observed at Chefe Donsa, followed by Sinana (8.2 t ha^{-1}), whereas the lowest mean grain yield was recorded at Holeta (4.4 t ha^{-1}) (Table 1). The highest mean performance of the genotypes for the thousand kernel weight (42.7 g) was attained at Chefe Donsa, whereas the lowest was found at Akaki (35.2 g). The pooled ANOVA over test environments indicated a significant ($p < 0.01$) impact of genotype, environment, and genotype by environment interactions on all traits evaluated (Supplementary Table 3). Furthermore, significant effects of replications and blocks were noted for traits, most likely due to variation within the field.

3.2 Variance components estimation, broad-sense heritability, and relationship between traits

The estimates of genotypic variance (σ_g^2) and genotypic coefficient of variation (GCV) for the thousand kernel weight (TKW) and grain-filling period (GFP) were high. The lowest σ_g^2 and GCV were obtained for grain yield (GYD) and the number of effective tillers (NET; Supplementary Table 4). The highest values of variance due to genotype by environment interaction (σ_{gxe}^2) and variance due to environments (σ_{gxe}^2) were recorded for plant height (PHT). In contrast, the number of effective tillers per plant (NET) showed the lowest values of both variances. The phenotypic coefficient of variation (PCV) ranged from 24.1 for days to maturity (DTM) to 137.5 for TKW. Most of the phenotypic traits evaluated in the present study showed high heritability (Supplementary Table 4). The highest broad sense heritability values were recorded for TKW ($H^2 = 0.97$) and GFP ($H^2 = 0.98$), indicating that these traits are highly heritable.

The Pearson correlation coefficients computed based on the BLUE mean values were positively significant ($p < 0.01$) for DTH with DTM, SPP, SPL, PHT and NET, for SPP with SPL, NET and GYD, and for SPL with PHT and NET (Figure 2). GYD was positively correlated with SPP ($r = 0.20$), and TKW ($r = 0.24$). Nevertheless, GYD had a negative correlation with DTH ($r = -0.22$) and PHT ($r = -0.38$) (Figure 2). DTH had a negative correlation with GFP and a positive correlation with DTM and SPP.

3.3 SNP markers distribution and density

In total, the 420 genotypes were genotyped with 24,145 SNP markers. Filtering of the genotypic data based on the number of

TABLE 1 Descriptive statistics for days to heading (DTH, in days), days to physiological maturity (DTM, in days), grain filling period (GFP, in days), plant height (PHT, in cm), spike length (SPL, in cm), grain yield (GYD, in t ha⁻¹), thousand kernel weight (TKW, in g), and the number of spikelets per spike (SPP, in counts) of 420 durum wheat genotypes grown in five test sites (ENV) in Ethiopia.

Traits	ENV	Mean	Median	Range	SE ^z	Traits	ENV	Mean	Median	Range	SE
DTH	Akaki	77.3	78	59–86	0.19	SPL	Akaki	6.24	6	4–12	0.04
	Chefe Donsa	74.8	76	68–84	0.14		Chefe Donsa	7.69	8	5–14	0.05
	Holeta	74.9	76	64–84	0.14		Holeta	8.16	8	5–13	0.06
	Kulumsa	68.0	68	59–77	0.13		Kulumsa	8.1	8	5–12	0.05
	Sinana	64.1	68	59–76	0.13		Sinana	9.48	10	4–14	0.07
	Pooled ENV	72.6	73	59–86	0.08		Pooled ENV	7.94	8	4–14	0.03
DTM	Akaki	138.5	137	133–153	0.12	GYD	Akaki	4.9	4.9	1.2–9.9	0.05
	Chefe Donsa	137.3	138	131–150	0.09		Chefe Donsa	8.4	8.5	1.6–13	0.06
	Holeta	132.6	133	127–147	0.09		Holeta	4.36	4.3	2.7–11	0.03
	Kulumsa	133.6	134	127–145	0.09		Kulumsa	7.78	7.8	1.8–13.5	0.06
	Sinana	138.1	138	130–150	0.08		Sinana	8.2	8.1	2.9–14	0.07
	Pooled ENV	136.0	136	127–153	0.06		Pooled ENV	6.74	6.7	1.2–14	0.03
GFP	Akaki	61.3	60	48–81	0.17	TKW	Akaki	35.25	35	20–51	0.16
	Chefe Donsa	62.5	62	51–77	0.13		Chefe Donsa	42.69	42	27–60	0.18
	Holeta	57.7	57	48–75	0.15		Holeta	42.21	42	29–60	0.17
	Kulumsa	65.6	65	53–80	0.16		Kulumsa	42.25	42	26–66	0.20
	Sinana	70.0	70	58–85	0.14		Sinana	42.05	41.5	21–57	0.18
	Pooled ENV	63.4	63	48–85	0.09		Pooled ENV	40.89	41	20–66	0.09
PHT	Akaki	69.5	69	43–111	0.31	SPP	Akaki	15.84	16	9–23	0.07
	Chefe Donsa	100.0	100	70–133	0.33		Chefe Donsa	18.14	18	13–23	0.06
	Holeta	93.4	94	57–129	0.40		Holeta	14.58	15	10–23	0.05
	Kulumsa	101.2	102	58–140	0.37		Kulumsa	18.14	18	13–24	0.06
	Sinana	117.0	117	65–156	0.40		Sinana	20.88	21	14–26	0.06
	Pooled ENV	96.3	98	43–156	0.28		Pooled ENV	17.52	18	9–26	0.04

^zSE, Standard error.

missing values, observed heterozygosity, and minor allele frequency resulted in 10,045 high quality polymorphic SNPs used for data analyses. Of these 10,045 SNPs, 4,807 (48%) and 5,238 (52%) were distributed on the A and B genomes, respectively (Table 2). The number of these SNPs per chromosome with regard to the two genomes ranged from 415 on chromosome 4B to 917 on chromosome 5B (Table 2).

The marker density was 1.01, 0.96, and 0.98 Mbp per marker for the A, B, and whole genomes, respectively. The SNP markers used in this study covered a total size of 9.86 Gbps, with chromosomes 1A and 2B having the smallest (584.2 Mbp) and largest (789.4 Mbp) regions (Table 2 and Figure 3).

3.4 Linkage disequilibrium

Among all possible pairs of SNPs on each chromosome, 490,775 pairs were found in LD (Table 3). Of these, 97,386 (19.8%) were found to be significant marker pairs with $r^2 \geq 0.3$ ($p < 0.01$; Table 3), which were therefore used to assess the MTAs. The significant marker pairs on each chromosome accounted for 12.1% ($r^2 = 0.12$ for chromosome 7A) to 26.2% ($r^2 = 0.211$ for chromosome 3B) of all marker pairs on the corresponding chromosomes (Table 3). The sudden LD decay among SNP pairs occurred within the range of 2.02–10.04 Mbp with an average of 4.26 Mbp (Supplementary Table 5;

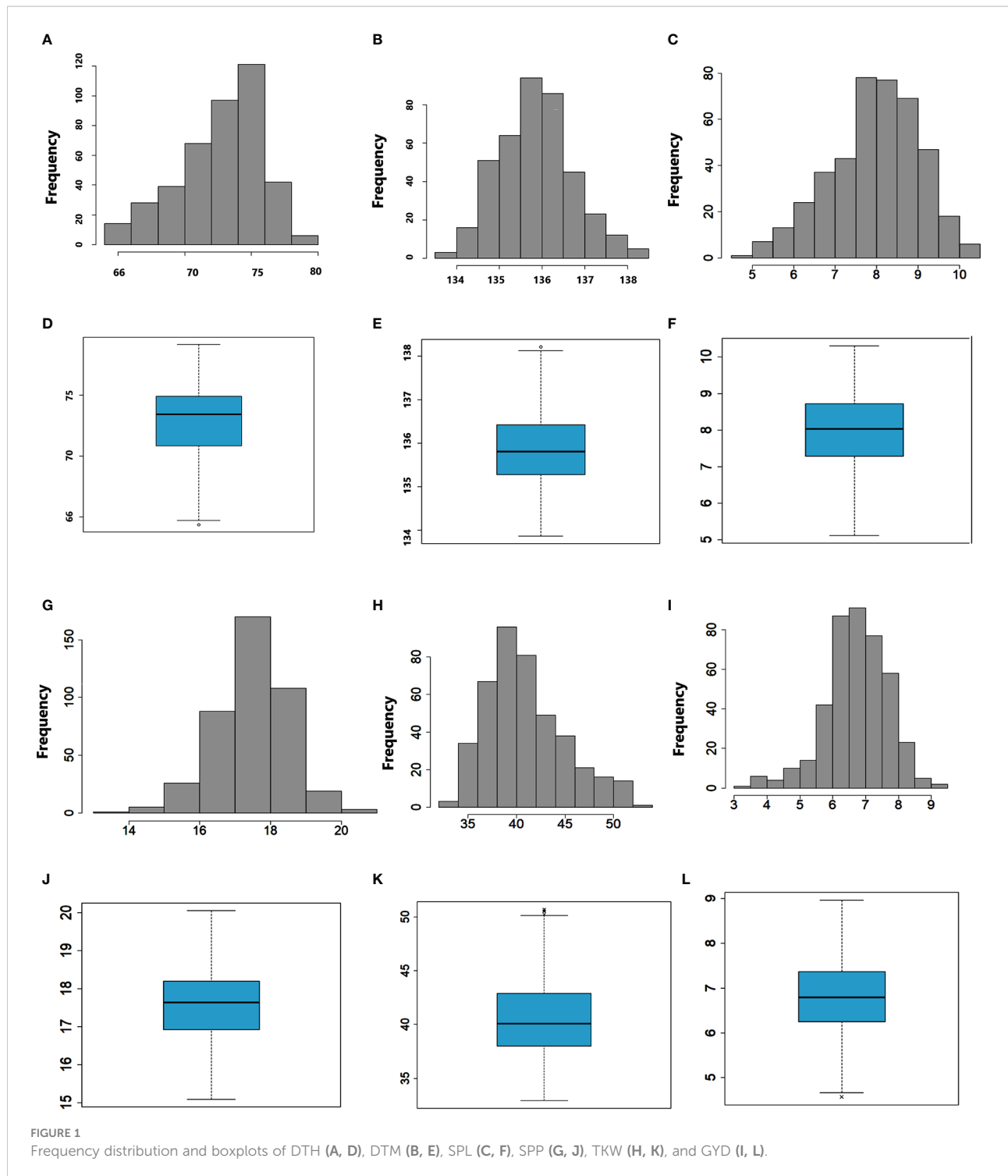


FIGURE 1
Frequency distribution and boxplots of DTH (A, D), DTM (B, E), SPL (C, F), SPP (G, J), TKW (H, K), and GYD (I, L).

Supplementary Figure 1). The fastest decrease of LD at cut-off ($r^2 = 0.3$) was observed on chromosome 7A. The r^2 values of marker pairs progressively declined as the physical distance between them increased on each chromosome (Supplementary Figure 1).

3.5 Principal component analysis, population structure, and kinship

The PCA scatter plot explained 92% (PC1 = 78.8% and PC2 = 13.2%) of the entire variation in the data set and grouped the

	DTH	DTM	SPP	SPL	PHT	NET	TKW	GFP
DTM	0.35 *							
SPP	0.44 *	-0.07 ^{ns}						
SPL	0.23 *	-0.01 ^{ns}	0.41 *					
PHT	0.18 *	0.35 *	0.08 ^{ns}	0.28 *				
NET	0.19 *	-0.08 ^{ns}	0.22 *	0.34 *	0.15 *			
TKW	-0.36 *	-0.01 ^{ns}	-0.13 *	-0.19 *	0.02 ^{ns}	-0.20 *		
GFP	-0.46 *	0.04 ^{ns}	-0.10 ^{ns}	0.05 ^{ns}	0.05 ^{ns}	-0.04 ^{ns}	0.13 *	
GYD	-0.22 *	-0.39 ^{ns}	0.20 *	0.06 ^{ns}	-0.38 *	0.05 ^{ns}	0.24 *	0.07 ^{ns}

FIGURE 2

Computed correlation plots between pairs of phenotypic traits based on the best linear unbiased estimators of the nine traits measured in 420 durum wheat genotypes. N.B. *refers to significant at $p < 0.01$, and ns refers to non-significant at $p < 0.01$.

genotypes into two subpopulations (Figure 4A). Subpopulation 1 contained almost all modern cultivars, and subpopulation two included all landraces by showing clear grouping based on genetic background (Figures 4A, B).

A model-based Bayesian cluster analysis using STRUCTURE revealed that the optimal uppermost clear true ΔK value was obtained at best when $K = 2$, suggesting the 420 genotypes form two subpopulations (Figure 4D). Based on this clustering, cluster-1 comprised 348 landraces and one cultivar (85.5% of the genotype), and cluster-2 comprised 33 cultivars and 28 landraces (14.5% of the

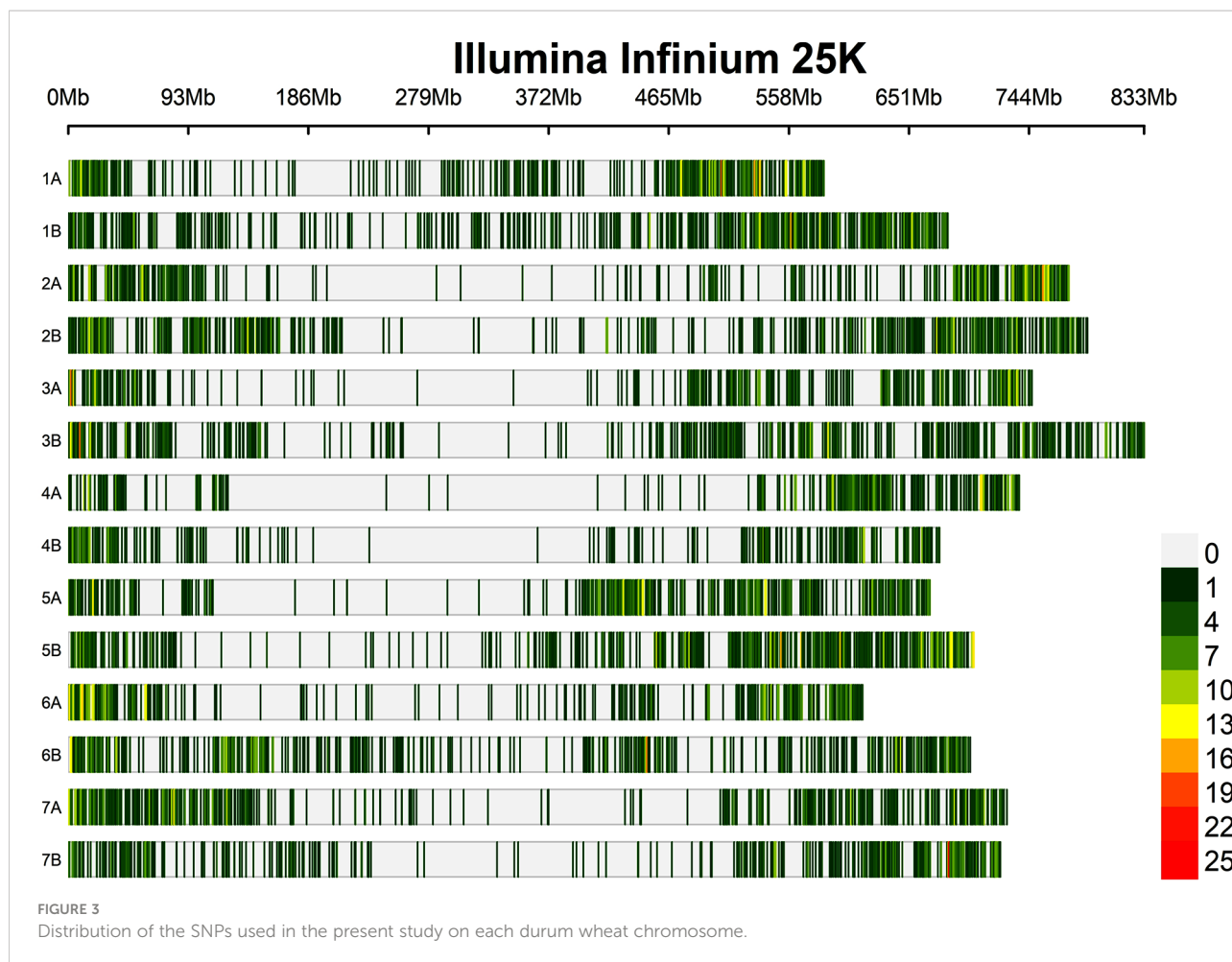
genotype). Analysis of admixture and purity using the STRUCTURE based on the Q value score ($Q < 0.80$ = admixture, and $Q > 0.80$ = pure genotypes) revealed that only 17 individuals (16 landraces and 1 cultivar) were classified as admixtures.

The STRUCTURE analysis revealed that 27 landraces were grouped with cultivars, and of these, three landraces (G242, G243, and G368) had a Q value of 1, which indicates 100% fitting the grouping with modern cultivars. Furthermore, one modern cultivar (G405), which was derived from related to landraces (according to their pedigree information), was grouped with landraces. The allelic

TABLE 2 The distribution of the 10,045 SNP markers across the entire durum wheat genome.

CHR ^z	NSPChr	GCR (bp)	SGRC (Mbp)	CHR	NSPChr	ROGC (bp)	SGRC (Mbp)
1A	757	1104472–585259074	584.2	1B	849	313555–681099620	680.8
2A	728	295475–774813964	774.5	2B	869	406084–789416853	789.4
3A	610	304055–746380464	746.1	3B	806	304239–746380464	746.1
4A	520	698412–736473645	735.8	4B	415	42526–674744571	674.7
5A	746	27537–667286510	667.3	5B	917	2555603–701346725	698.8
6A	643	591650–615260837	614.7	6B	753	2052283–698554772	696.5
7A	803	171878–727023089	726.9	7B	629	47368–721753586	721.7
A ^a	4807	na	4849.5	B ^b	5238	na	5008

^z CHR, chromosome; ^aA, A genome; ^bB, B genome; NSPChr, Number of SNPs per chromosome; GCR, Genome coverage range; SGRC, Size of genomic region(s) covered; na, Not applicable.



divergence between the two subpopulations inferred by STRUCTURE was 0.27. On average, the expected heterozygosity of subpopulation-1 (CI - I) and subpopulation-2 (CI - II) was 0.22 and 0.32, respectively. Subpopulation-1 had a mean F_{ST} value of 0.62, while subpopulation-2 had a mean F_{ST} value of 0.35, indicating high differentiation among the individuals of each population. Although a slight difference was observed, model-based Bayesian clustering and distance-based PCA similarly grouped the individuals into two subpopulations. The kinship matrix heatmap revealed familial relationships between the genotypes, which can be regarded as intermediate on average (Figure 4C).

3.6 GWAS scan of phenotypic traits

Considering all test locations and combined data over locations, GWAS was able to identify 179 significant MTAs for the nine traits (Supplementary Table 6). Of these, 23 MTAs were detected for DTH, 32 MTAs for DTM, 15 MTAs for GFP, 8

MTAs for GYD, 5 MTAs for NET, 19 MTAs for PHT, 26 MTAs for spike length (SPL), 12 MTAs for SPP and 39 MTAs for TKW. Using BLUES of combined data across the five environments revealed 44 significant MTAs for the nine traits evaluated in this study. Further results and discussions (below) focus on these significant MTAs identified using the combined data across the five environments. The Manhattan and quantile-quantile (Q-Q) plots for each trait and environment are presented in Supplementary Figures 2A-E, respectively.

3.6.1 Marker trait association for phenological traits

For phenological traits (DTH, DTM, and GFP), 12 significant MTAs were identified from the GWAS of combined data from the five locations (Table 4). The GWAS scan for DTH detected six significant MTAs on chromosomes 1B, 2A, 5B, 6B, 7A, and 7B (Figure 5 and Table 4). The Q-Q plot showed that the data fitted the model well, and false positive MTAs were controlled. Among these MTAs, three were previously reported (Golabadi et al., 2011;

TABLE 3 A summary of linkage disequilibrium analysis for SNP marker pairs and the distribution of significant SNP pairs across each chromosome of each genome.

Chromosome	Total number of SNP pairs	Significant SNP marker pairs at $r^2 \geq 0.3$ ($p < 0.01$)	Average r^2	Average distance (Mbp ^z)
1A	35,875	7,811 (21.8%)	0.20	20.4
1B	41,700	7,666 (18.4%)	0.16	21.1
2A	35,700	7,218 (20.2%)	0.17	28.3
2B	42,700	8,709 (20.4%)	0.17	23.8
3A	29,800	5,297 (17.8%)	0.16	32.9
3B	39,550	10,362 (26.2%)	0.21	27.4
4A	25,250	3,580 (14.2%)	0.14	38.4
4B	20,000	3,805 (19.0%)	0.17	44.9
5A	36,600	8,144 (22.3%)	0.19	23.7
5B	45,100	8,738 (19.4%)	0.17	20.1
6A	31,450	8,454 (26.9%)	0.22	25.6
6B	36,900	7,586 (20.6%)	0.18	24.6
7A	39,450	4,781 (12.1%)	0.12	23.9
7B	30,700	5,234 (17.0%)	0.15	30.8
A genome	234,125	45,284 (19.3%)	0.17	26.9
B genome	293,250	52,102 (17.8%)	0.17	25.8
Total	490,775	97,386 (19.8%)	0.17	26.8

^z Mbp, Mega base pair.

Giraldo et al., 2016; Mangini et al., 2018), while three MTAs on the B genome (AX-109859693, wsnp_BE496986B-Ta_2_2, Ku_c24482_1132) were novel. The significant MTAs explained 1.1 to 22.2% of the total phenotypic variation in DTH. Among the significant MTAs for DTH, wsnp_BE496986B-Ta_2_2, AX-109859693, Ku_c24482_1132, and IACX11338 appeared significant in two or more test environments and hence can be regarded as stable MTAs.

The GWAS scan detected four significant MTAs across test environments for DTM. Of these MTAs, Kukri_rep_c73477_888 on chromosome 6A was previously reported (Mangini et al., 2018) and was detected in two environments. Three MTAs (Tdurum_contig49186_437 and Tdurum_contig12722_779 on chromosome 7A; and AX-109869840 on chromosome 6A) were likely to be potential new loci (Table 4 and Figure 5). The proportion of phenotypic variance explained by these four significant SNPs ranged from 2 to 33%.

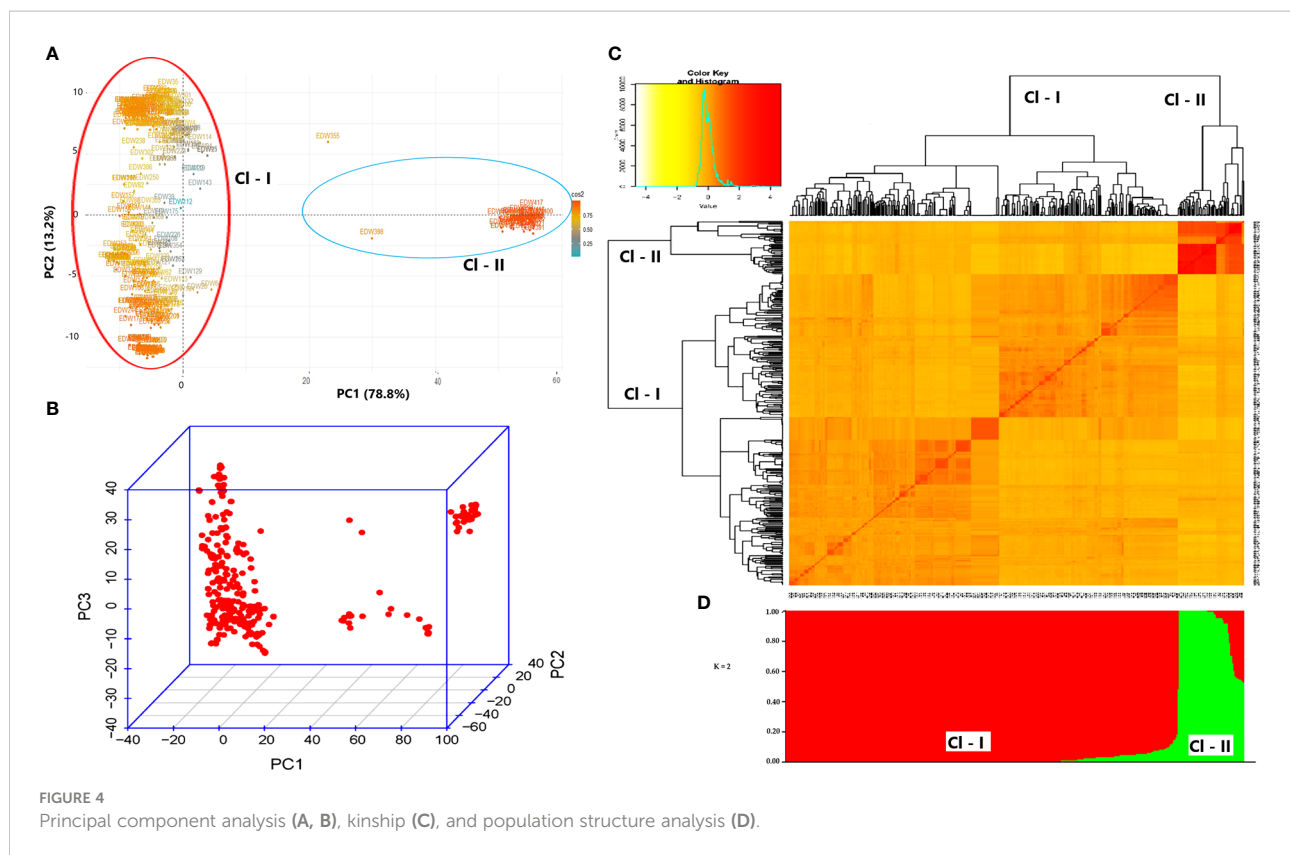
GWAS revealed two significant MTAs for GFP on chromosomes 3B and 7B (Table 4 and Figure 5). The RAC875_c62223_86 MTA on 3B was previously reported (Giraldo et al., 2016). However, Kukri_c60966_261 on chromosome 7B was novel and detected repeatedly in two locations. RAC875_c62223_86 and Kukri_c60966_261 explained

4% and 31% of the variation in GFP obtained in the present study, respectively.

3.6.2 Marker trait association for plant architecture

The GWAS analysis revealed 14 MTAs significantly associated with plant architecture traits (Table 4 and Figure 6). For PHT, six significant MTAs were detected (Figure 6). Among these, four (AX-158602974, and AX-95259256 on chromosome 1B, wsnp_BE443745A-Ta_2 on chromosome 5A, and BS00091519_51 on chromosome 5B) were previously reported (Zhang et al., 2012; Mengistu et al., 2016; Roncallo et al., 2017). The other two MTAs (AX-95154560 and Tdurum_contig75127_589 on chromosomes 1B and 7B, respectively) were novel. The six significant MTAs explained 1.2 to 23% of the variation in PHT recorded in this study.

The MTA analysis for spike length (SPL) revealed five significant MTAs. Among the significant MTAs, Tdurum_contig45715_1246 on chromosome 1B was previously identified (Giraldo et al., 2016). The remaining four MTAs (Kukri_c17062_618 and Tdurum_contig76960_213 on chromosome 2A, Kukri_c3096_1411 on chromosome 2B, and AX-94615777 on chromosome 5A) are novel (Figure 6). The



proportion of the variation in SPL elucidated by the significant MTAs varied from 2.75% to 12.3%. For NET, GWAS revealed two (GENE-0410_71 and AX-94782013) significant MTAs on chromosome 1B and 7B, respectively (Table 4 and Figure 6), which have not been previously reported. These MTAs explained 2.8% and 5.6% of the variation in NET, respectively.

3.6.3 Marker trait association for grain yield and related traits

GWAS for grain yield and yield-related traits evidenced 18 significant MTAs (Table 6 and Figure 7). The association scan for SPP resulted in seven significant MTAs on chromosomes 1B, 2B, 3A, 4A, and 7A. The proportion of phenotypic variance explained by the associated MTAs ranged from 1.1% to 17%. Of the seven significant MTAs for SPP, four (AX-89760660, Tdurum_contig25602_212, BS00110281_51, and AX-158591111) were previously reported (Golabadi et al., 2011; Mengistu et al., 2016; Soriano et al., 2016; Kidane et al., 2017a; Roncallo et al., 2017; Abu-Zaitoun et al., 2018; Mangini et al., 2018), whereas the remaining three (RAC875_c400_193, AX-158597411, and AX-94631122) SNPs are novel.

For GYD, GWAS revealed four significant MTAs on chromosomes 1B, 5A, 5B, and 7A. The proportion of phenotypic variance explained by the significant MTAs ranged

from 1.74% (RAC875_c57656_170 on chromosome 7A) to 44.95% (IAAV3365 on chromosome 5A). Alleles of the high signal MTAs (locus IAAV3365, A/G alleles) had a highly significant effect on grain yield (Figure 8A). The genotypes carrying allele A had higher average grain yield across the five environments as compared to genotypes carrying allele G. RAC875_c57656_170 was previously reported for GYD (Maccaferri et al., 2014), whereas the remaining three MTAs (IAAV3365, RFL_Contig3481_1669 on chromosome 1B, and Excalibur_c51720_84 on chromosome 5B) were newly detected in the present study.

The genome-wide association analysis identified seven significant MTAs for TKW on chromosomes 1B, 3B, 5A, 6A, and 7A. The phenotypic variance explained by the associated SNPs ranged from 1.05% to 10.6%. Among the MTAs significantly associated with TKW, two (AX-158606713 and wsnp_Ex_rep_c66939_65371026 on chromosomes 1B and 7A, respectively) were previously reported. However, the other five MTAs (BS00071597_51, AX-158541767, RAC875_c41315_157, AX-158564275, and AX-94640059 on chromosomes 3B, 5A, 6A, and 7A, respectively) were novel. The effect of alleles on locus AX-158564275 (A/G alleles) revealed a highly significant difference in TKW (Figure 8B). The genotypes with the allele A had high TKW compared to genotypes carrying allele G.

TABLE 4 Summary of significant marker-trait associations for the nine traits revealed based on the combined data of the five locations on each durum wheat chromosome (CHR).

SNP (MTAs)	CHR	POS (bp)	P-value	MAF	Effect	PVE ^Z	Trait
AX-158591262	7A	30075367	5.13281E-07	0.31	-0.51	1.62	DTH
AX-94884567	2A	756029643	5.54677E-08	0.16	0.77	1.1	
IACX11338	1B	522669860	1.43456E-06	0.09	-2.04	22.18	
AX-109859693	5B	5172617	3.78017E-09	0.05	1.16	5.13	
w SNP_BE496986B_Ta_2_2	6B	568039035	1.80668E-08	0.17	-0.68	5.28	
Ku_c24482_1132	7B	155935903	3.69626E-07	0.29	-0.66	1.67	
AX-109869840	6A	603461435	2.75361E-06	0.05	-0.74	21.07	DTM
Kukri_rep_c73477_888	2A	70572673	1.55589E-23	0.09	-2.54	32.88	
Tdurum_contig49186_437	7A	32558067	5.98292E-09	0.49	0.29	1.88	
Tdurum_contig12722_779	7A	44540113	4.95657E-07	0.07	0.41	4.45	GFP
RAC875_c62223_86	3B	763192473	1.31126E-07	0.32	-0.53	3.58	
Kukri_c60966_261	6B	693337622	5.50108E-07	0.08	-1.41	30.67	GYD
RFL_Contig3481_1669	1B	4043159	5.87000E-06	0.10	0.02	5.45	
Excalibur_c51720_84	7A	709197555	6.71000E-06	0.05	0.02	15.75	
RAC875_c57656_170	7A	614197852	1.75039E-06	0.46	0.17	1.74	
IAAV3365	5A	548344620	1.87701E-10	0.06	-0.79	44.95	NET
GENE-0410_71	1B	523053033	1.78840E-07	0.44	0.22	5.61	
AX-94782013	7B	604310198	3.10157E-06	0.09	-0.25	2.79	PHT
AX-158602974	1B	580658793	5.16606E-11	0.06	-4.33	2.68	
BS00091519_51	5B	5174649	2.79112E-06	0.05	-2.03	9.64	
AX-158521163	1B	669849432	3.25148E-06	0.06	-2.10	1.2	
AX-95259256	1B	629504430	2.07679E-07	0.12	-1.52	4.51	
w SNP_BE443745A_Ta_2_1	5A	439542987	7.51027E-08	0.37	1.12	2.59	
AX-95154560	6B	120830636	5.51278E-08	0.05	2.73	23.76	
Tdurum_contig75127_589	7B	697951769	5.09222E-08	0.06	-3.88	8.34	SPL
Tdurum_contig45715_1246	1B	314321199	6.74631E-07	0.48	-0.26	3.62	
Kukri_c17062_618	2A	522595273	4.96382E-07	0.28	0.17	2.75	
Tdurum_contig76960_213	2A	492195805	3.17135E-07	0.13	-0.63	12.3	
Kukri_c3096_1411	2B	314134332	3.40111E-07	0.22	0.28	3.02	
AX-94615777	5A	529858969	2.51717E-07	0.45	0.15	1.01	SPP
Tdurum_contig25602_212	2B	546442999	3.25386E-07	0.24	0.22	1.76	
AX-158591111	7A	33518205	7.03418E-08	0.24	-0.24	1.85	
BS00110281_51	4A	724872914	1.51333E-06	0.06	0.41	4.5	
AX-89760660	1B	519060573	8.65163E-09	0.07	-0.49	16.82	
RAC875_c400_193	1B	1547605	8.21951E-07	0.10	-0.25	1.53	
AX-158597411	2B	99223728	2.38152E-06	0.49	0.24	2.27	
AX-94631122	3A	723577013	1.23025E-09	0.23	0.31	1.12	

(Continued)

TABLE 4 Continued

SNP (MTAs)	CHR	POS (bp)	P-value	MAF	Effect	PVE ^Z	Trait
AX-158606713	1B	546979073	1.26545E-06	0.26	-0.67	1.05	TKW
w SNP_ Ex_ rep_ c66939_65371026	7A	6480158	1.99623E-06	0.18	-0.81	1.72	
BS00071597_51	3B	803879943	7.43197E-09	0.20	1.44	7.93	
AX-158541767	3B	61267921	2.83169E-07	0.07	-1.82	10.6	
RAC875_c41315_157	5A	431829169	3.78422E-06	0.10	-1.28	5.65	
AX-158564275	6A	528989018	3.23204E-10	0.09	-1.57	5.45	
AX-94640059	7A	686968079	4.86263E-06	0.44	0.65	1.23	

Chr, Chromosome; POS, Physical position of SNP; bp, Base pair; MAF, Minor allele frequency; PVE, Phenotypic variance explained; DTM, Days to heading; DTM, Days to maturity; GFP, Grain filling period; NET, Number of effect tillers per plant; SPL, Spike length; PHT, Plant height; SPP, Number of spikelets per spike; TKW, Thousand kernel weight; GYD, Grain yield.

3.6.4 Identification of putative novel MTAs and their underlying candidate genes

According to the LD decay information for each chromosome, a genomic region of ten Mbp around each significant SNP (five Mbp downstream and five Mbp upstream of the significant SNP) is considered to be a QTL. Significant SNPs within the ranges of 10 Mbp apart are considered to refer to the same QTL. Based on this approach, 37 QTLs were identified for the 44 significant MTAs (Table 5). The names of these QTLs (*q.gwas.01* to *q.gwas.37*) are provided in the first column of Table 5. Among the 37 QTLs, 16 were located in or near genomic regions previously reported for the corresponding traits (Golabadi et al., 2011; Maccaferri et al., 2014; Giraldo et al., 2016; Mengistu et al., 2016; Soriano et al., 2016; Kidane et al., 2017a; Abu-Zaitoun et al., 2018; Mangini et al., 2018; Roncallo et al., 2018), while 21 were novel (Table 5). Two QTLs for SPP (*q.gwas.15* and *q.gwas.20*) and one QTL (*q.gwas.08*) for TKW were previously described based on Ethiopian durum wheat germplasm (Mengistu et al., 2016; Kidane et al., 2017a). Genomic regions for five putative QTLs (*q.gwas.01*, *q.gwas.02*, *q.gwas.22*, *q.gwas.24*, and *q.gwas.30*) overlap with more than one trait evaluated in this study (Table 5). For example, four significant MTAs for DTH, DTM, SPP, and TKW were co-localized and hence were considered to be referring to the same QTL (*q.gwas.30*) (Table 5). The analysis of the sequences of these putative QTLs genomic regions based on durum wheat, the reference genome at the Ensemble Plants database, led to the identification of 774 potential candidate genes (Supplementary Table 7).

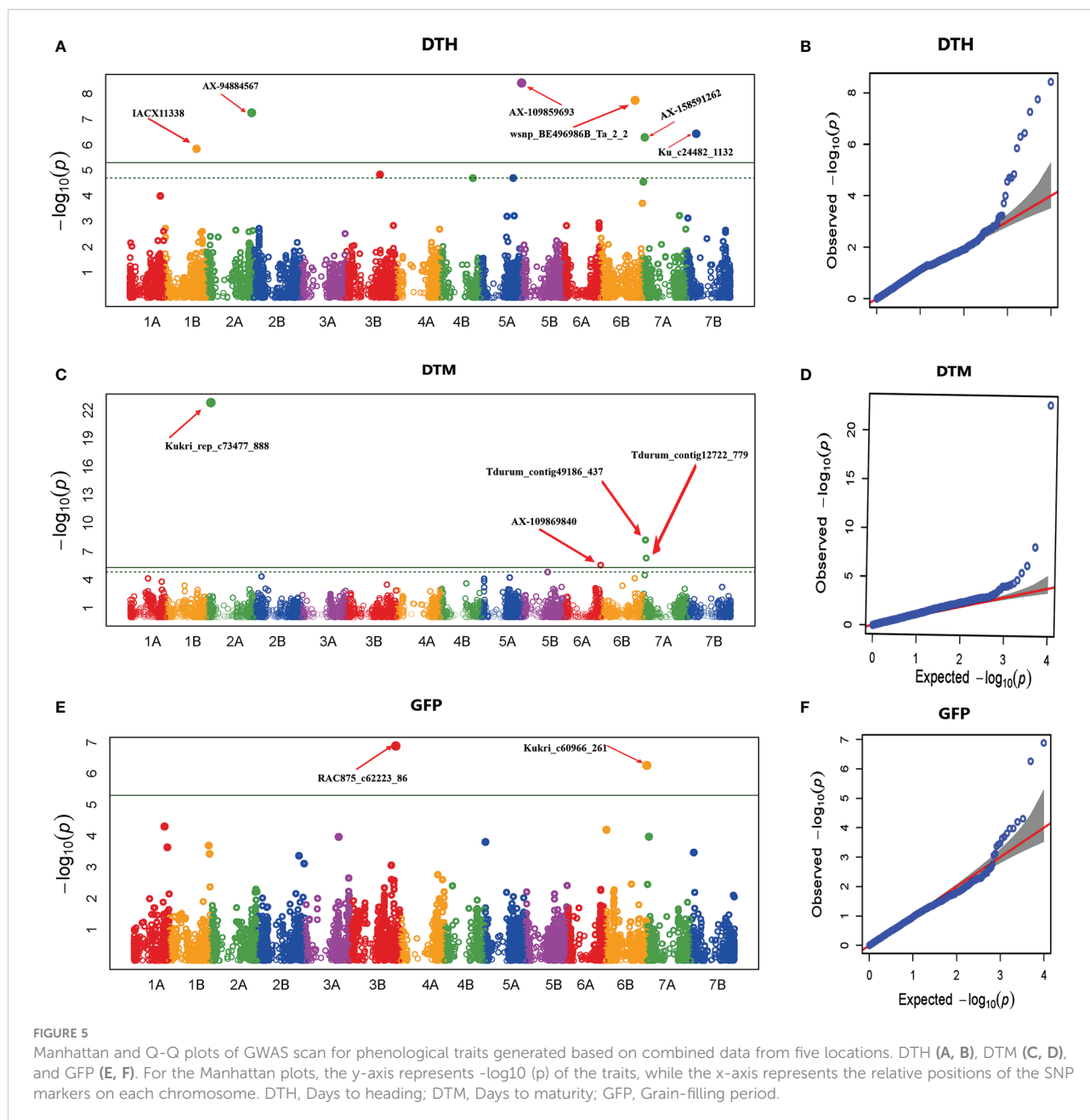
The significance of the candidate genes was evaluated by reviewing previously published genomic regions associated with the traits targeted in the present study (Zhang et al., 2012; Maccaferri et al., 2016; Maccaferri et al., 2019; Kidane et al., 2019; Mazzucotelli et al., 2020; Zhao et al., 2020). This resulted in 32 genes related to eight of the nine target traits in durum wheat (Table 6). The putative candidate genes *TRITD7Av1G01175*, *TRITD7Av1G017240*, and *TRITD7Av1G017550* (all located on chromosome 7A), which encode Growth-regulating factor, Zinc-finger CCCH domain protein TE, and NAC domain-containing

protein, respectively, were associated with DTH and DTM. The genes *TRITD1Bv1G168480* and *TRITD7Bv1G057630* encode WRKY transcription factor and Flowering Locus T/Terminal Flower 1-like protein, respectively, were reported to regulate DTH. The *TRITD1Bv1G000110* gene, on chromosome 1B, which encodes Tryptophan aminotransferase-related protein 2, was shown to be associated with SPP and GYD (Table 6).

4 Discussion

This study used GWAS to define durum wheat genomic regions associated with phenological, plant architecture, grain yield, and yield-related traits. Furthermore, analyses of population structure and linkage disequilibrium were carried out to increase the efficiency of detecting reliable marker-trait associations as well as identifying the genetic basis of those associations. The present study utilized a large number of diverse durum wheat landraces and cultivars, which were grown across diverse environments in Ethiopia. This facilitated the identification of novel SNP loci associated with nine durum wheat phenotypic traits, including grain yield and grain yield-related traits. The present study findings have significant implications for both the development of molecular markers for genomics-led breeding and for providing new insights into the architecture of genomic regions regulating various traits of interest in durum wheat. These could facilitate the improvement of grain yield and other desirable characteristics to support global food security.

To meet the growing demand for durum wheat grains as well as the challenges to its production brought about by the expanding environmental changes, it is imperative that the genetic resources of durum wheat, including landraces, modern cultivars, and breeding lines, be effectively utilized for breeding new cultivars (Kankwatsa et al., 2017; Maccaferri et al., 2019b; Kumar et al., 2020; Mazzucotelli et al., 2020). As such, identifying loci that regulate desirable traits in breeding

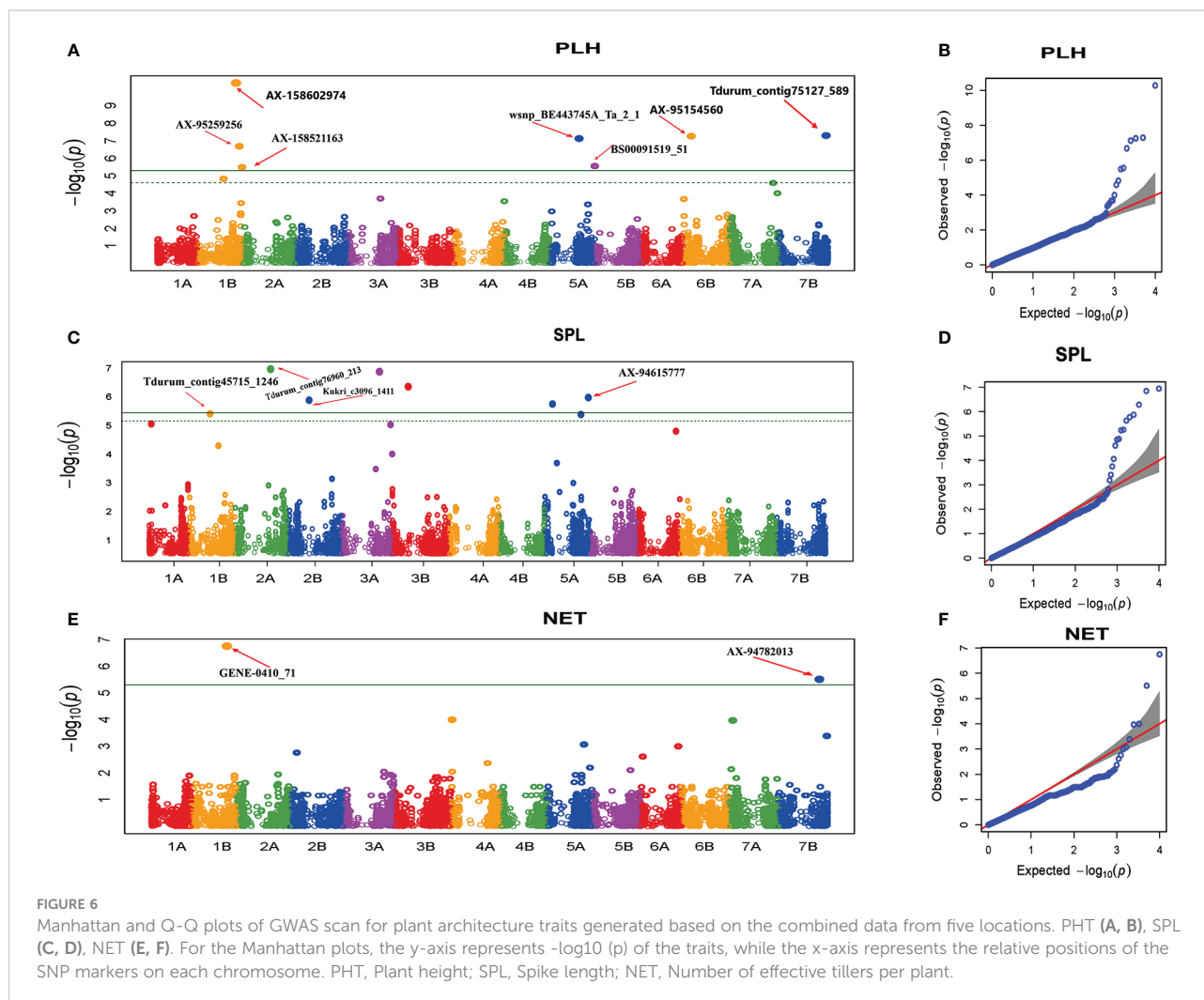


programs helps to develop markers for marker-assisted breeding, thus contributing to food security (Garcia et al., 2019; Wang et al., 2019; Mérida-García et al., 2020).

The present study revealed highly significant contributions of genotypes, environments, and genotype by environment interactions to the phenotypic variations of the target traits ($p < 0.001$), which is consistent with the results of previous research on durum wheat (Mengistu et al., 2015; Mohammadi et al., 2018; Mekonnen et al., 2021). The observed high genotypic variance, genotypic coefficient of variation, and broad-sense heritability for TKW and GFP, strongly suggest that their variation is mainly due to heritable genetic differences among

the landraces and cultivars. There was a low genotypic variance and genotypic coefficient of variation for GYD, indicating the challenges associated with improving this trait. Nevertheless, moderate to a high level of broad-sense heritability were recorded for all traits, meaning that a significant part of the observed variation is heritable and that the results agree with previous findings in durum wheat (Sukumaran et al., 2018; Alemu et al., 2020a).

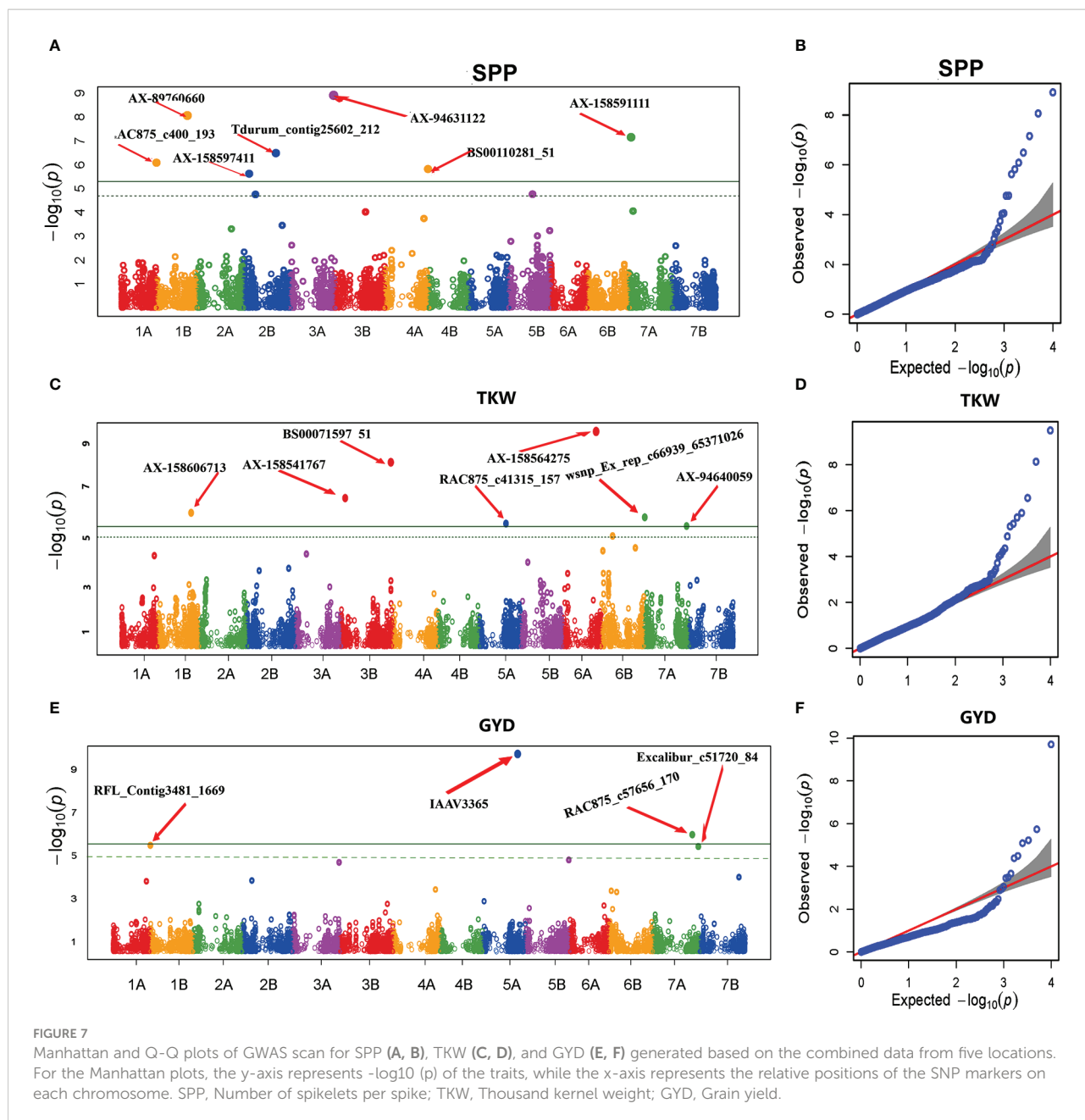
The present study found that GYD had a moderate but significant ($p < 0.01$) positive correlation with SPP, and TKW, indicating that the simultaneous selection of desirable characteristics of these traits could lead to the improvement of



grain yield in this crop. However, GYD negatively correlated with DTH and PHT, indicating that late-heading genotypes generally have lower grain yield than early-heading types. However, the early-heading types appear to have a more extended grain-filling period, as a very low but significant positive correlation was obtained for GYD versus DTM. TKW exhibited a moderate positive correlation with GYD, and GFP, implying that direct improvement of these traits may improve the former, which contributes to enhancing GYD. Conversely, TKW had a negative relation with DTH and DTM. Thus late-maturing cultivars will have a relatively low TKW.

The LD among the SNP marker pairs showed a sharp decline within the physical distance ranging from 2.02 to 10.4 Mbp, with an average of 4.26 Mbp. This decline in LD is far below the results of previous research using Ethiopian durum wheat landraces (Alemu et al., 2021b), which reported an average physical distance of 69.1 Mbp. Similarly, Mekonnen et al., (2021) found a higher mean LD decay (31.44 Mbp, $r^2 = 0.2$)

in their study on diverse Ethiopian bread wheat germplasm. This disparity could arise due to the type and density of markers, genomic regions the markers cover, and differences in the sample used in these studies. However, Fayaz et al., 2019 found a low LD decay (2–3 cM) of the A and B sub-genomes using Iranian durum wheat landraces at a critical $r^2 = 0.11$. Likewise, Rufo et al., (2019) noted an LD decay ranging from 1 to 9 cM on A and B genomes from landraces and released cultivars of Mediterranean wheat. The fastest LD decay rate of an average physical distance of 2.02 Mbp was recorded for chromosome 7A. In contrast, the slowest was recorded for chromosome 4A (10.04 Mbp), which indicates the differences in recombination rates among different genomic regions of different chromosomes. On average, the A genome showed a more rapid LD decay than the B genome (Supplementary Table 4), and more substantial selection pressure could be partly caused in the A genome than in the B genome (Liu et al., 2019; Kumar et al., 2020). This result most likely confirms the impact of



genetic drift, mutation, gene flow, recombination, the pressure of population selection, and historical events on both A and B genomes (Fayaz et al., 2019).

The population clustering inferred by STRUCTURE and PCA divided the genotypes into two sub-populations, similar to the results of earlier research (Wang et al., 2019; Alemu et al., 2020b; Kumar et al., 2020; Mekonnen et al., 2021). Based on Q-score values of STRUCTURE analysis ($Q > 0.80$), 96% of the landraces were pure, and 4% of the genotype were admixtures. The kinship matrix was used to estimate the family relatedness and to confirm the relation within the genotypes. Hence, the cumulative results from STRUCTURE, PCA, and kinship

suggest adjusting the GWAS model to avoid bias arising from spurious associations, thereby reducing false-positive associations arising from co-ancestry. Moreover, FarmCPU, a robust statistical model for GWAS, adequately accounted for the spurious associations that arose from population structure, cryptic relatedness, and marker effects, as shown by Q-Q plots. Based on the five-test sites' mean data, the GWAS revealed 44 MTAs. The SNPs associated with the target traits were distributed across the whole chromosome except chromosome 1A, which did not bear any significant MTAs.

Using GWAS, different genomic regions associated with grain yield were identified in the present study. The putative

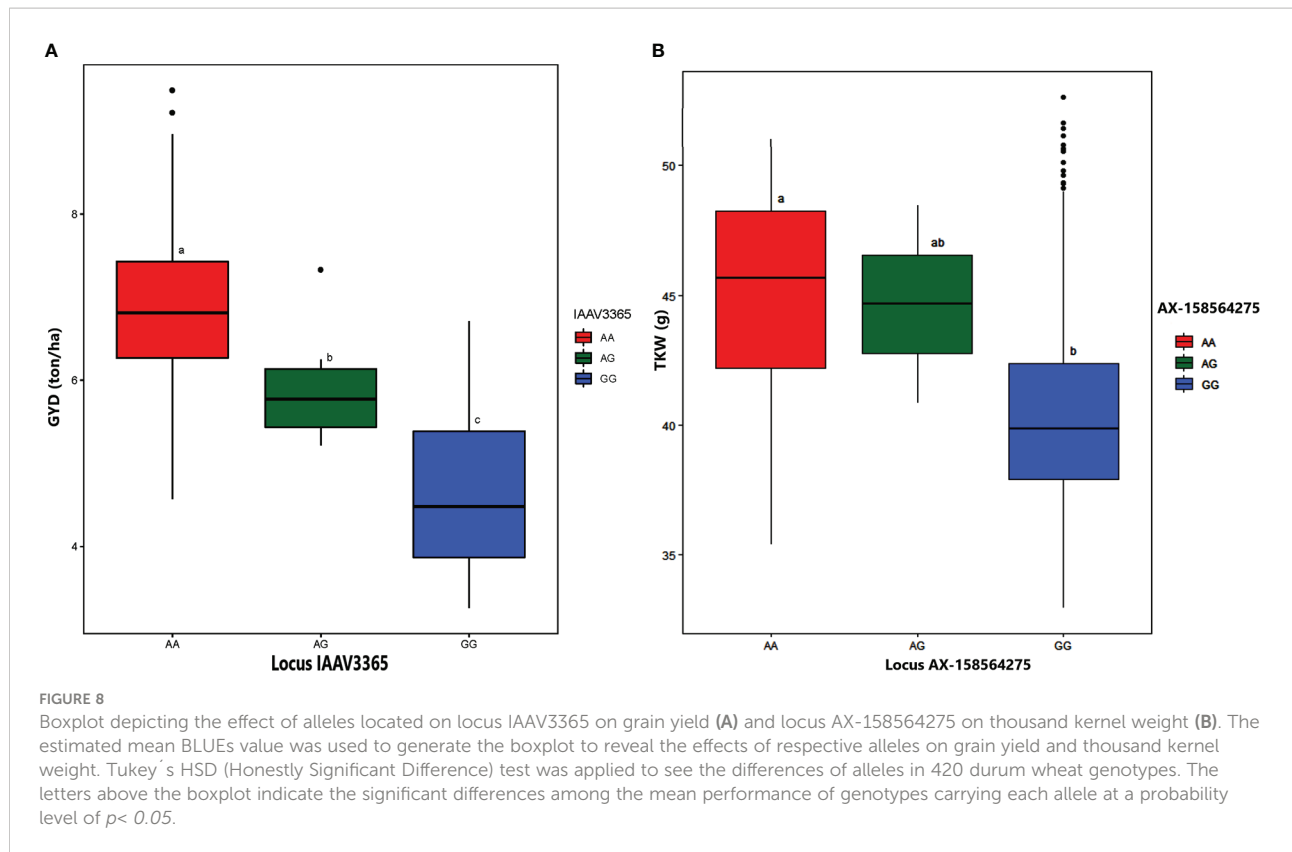


TABLE 5 Summary of putative quantitative trait loci (QTLs) identified for the nine phenotypic traits analyzed in the present study using Ethiopian durum wheat germplasm.

Putative QTL	Associated SNP	Chr	SNP position (bp)	TAPQTL
<i>q.gwas.01^a</i>	IACX11338	1B	522669860	DTH
<i>q.gwas.02^a</i>	RFL_Contig3481_1669	1B	4043159	GYD
<i>q.gwas.01^a</i>	GENE-0410_71	1B	523053033	NET
<i>q.gwas.03^a</i>	AX-158602974	1B	580658793	PHT
<i>q.gwas.04^a</i>	AX-158521163	1B	669849432	PHT
<i>q.gwas.05</i>	AX-95259256	1B	629504430	PHT
<i>q.gwas.06^a</i>	Tdurum_contig45715_1246	1B	314321199	SPL
<i>q.gwas.02</i>	RAC875_c400_193	1B	1547605	SPP
<i>q.gwas.07</i>	AX-89760660	1B	519060573	SPP
<i>q.gwas.08^a</i>	AX-158606713	1B	546979073	TKW
<i>q.gwas.09^a</i>	AX-94884567	2A	756029643	DTH
<i>q.gwas.10</i>	Kukri_rep_c73477_888	2A	70572673	DTM
<i>q.gwas.11</i>	Tdurum_contig76960_213	2A	492195805	SPL
<i>q.gwas.12</i>	Kukri_c17062_618	2A	522595273	SPL
<i>q.gwas.13</i>	Kukri_c3096_1411	2B	314134332	SPL

(Continued)

TABLE 5 Continued

Putative QTL	Associated SNP	Chr	SNP position (bp)	TAPQTL
<i>q.gwas.14</i> ^a	AX-158597411	2B	99223728	SPP
<i>q.gwas.15</i>	Tdurum_contig25602_212	2B	546442999	SPP
<i>q.gwas.16</i>	AX-94631122	3A	723577013	SPP
<i>q.gwas.17</i> ^a	RAC875_c62223_86	3B	763192473	GFP
<i>q.gwas.18</i>	AX-158541767	3B	61267921	TKW
<i>q.gwas.19</i>	BS00071597_51	3B	803879943	TKW
<i>q.gwas.20</i> ^a	BS00110281_51	4A	724872914	SPP
<i>q.gwas.21</i>	IAAV3365	5A	548344620	GYD
<i>q.gwas.22</i>	wsnp_BE443745A_Ta_2_1	5A	439542987	PHT
<i>q.gwas.23</i>	AX-94615777	5A	529858969	SPL
<i>q.gwas.22</i>	RAC875_c41315_157	5A	431829169	TKW
<i>q.gwas.24</i> ^a	AX-109859693	5B	5172617	DTH
<i>q.gwas.24</i> ^a	BS00091519_51	5B	5174649	PHT
<i>q.gwas.25</i> ^a	AX-109869840	6A	603461435	DTM
<i>q.gwas.26</i>	AX-158564275	6A	528989018	TKW
<i>q.gwas.27</i>	wsnp_BE496986B_Ta_2_2	6B	568039035	DTH
<i>q.gwas.28</i>	Kukri_c60966_261	6B	693337622	GFP
<i>q.gwas.29</i>	AX-95154560	6B	120830636	PHT
<i>q.gwas.30</i> ^a	AX-158591262	7A	30075367	DTH
<i>q.gwas.30</i> ^a	Tdurum_contig49186_437	7A	32558067	DTM
<i>q.gwas.31</i>	Tdurum_contig12722_779	7A	44540113	DTM
<i>q.gwas.32</i>	RAC875_c57656_170	7A	614197852	GYD
<i>q.gwas.33</i> ^a	wsnp_Ex_c16045_24471413	5B	685974689	GYD
<i>q.gwas.30</i> ^a	AX-158591111	7A	33518205	SPP
<i>q.gwas.30</i> ^a	wsnp_Ex_rep_c66939_65371026	7A	6480158	TKW
<i>q.gwas.34</i>	AX-94640059	7A	686968079	TKW
<i>q.gwas.35</i>	Ku_c24482_1132	7B	155935903	DTH
<i>q.gwas.36</i>	AX-94782013	7B	604310198	NET
<i>q.gwas.37</i>	Tdurum_contig75127_589	7B	697951769	PHT

^aPreviously identified QTLs. Chr, chromosome; DTH, Days to heading; DTM, Days to physiological maturity; GFP, Grain filling; NET, Number of effective tillers per plant; GFP, Grain filling period; PHT, Plant height; SPL, Spike length; SPP, Number of spikelets per spike; TKW, Thousand kernel weight; GYD, Grain yield; TAPQTL, Traits associated with Putative QTL.

QTLs identified for this trait are *q.gwas.02* (0.3 – 5.8 Mbp) on chromosome 1B, *q.gwas.21* (544.7 – 554.9 Mbp) on chromosome 5A, *q.gwas.32* (609.2 – 619.2 Mbp) on chromosome 7A, and *q.gwas.33* (703.3 – 714.2 Mbp) on chromosome 7A. Among them, *q.gwas.02*, *q.gwas.21*, and *q.gwas.32* are novel QTLs, as these genomic regions have not previously been reported for their association with grain yield. The putative QTL *q.gwas.33* on chromosome 5B is co-localized

within the same genomic region of a QTL reported by [Maccaferri et al., \(2014\)](#); [Maccaferri et al., \(2016\)](#) for grain yield and total root numbers, respectively using durum wheat recombinant inbred lines. The QTL regions of *q.gwas.33* is also identified for spikes per plant ([Mengistu et al., 2016](#)), kernel Fe content ([Velu et al., 2017](#)), kernels per spikelets ([Peng et al., 2003](#)), fusarium head blight resistance ([Ghavami et al., 2011](#)), yellow rust resistance ([Liu et al., 2017a](#)), and stem rust resistance

TABLE 6 Summary of selected genes associated with some of the putative QTLs identified in the present study.

S/N	SNP	Chr	PQAG	Gene ID	GSS	GSE	Description of gene	Trait
1	IACX11338	1B	<i>q.gwas.01</i>	TRITD1Bv1G168480	523422292	523424581	WRKY transcription factor	
2	IACX11338	1B	<i>q.gwas.01</i>	TRITD1Bv1G169970	526873942	526876463	WRKY transcription factor	
3	IACX11338	1B	<i>q.gwas.01</i>	TRITD1Bv1G170060	527017659	527018012	WRKY DNA-binding protein 39 G	
4	AX-158591262	7A	<i>q.gwas.30</i>	TRITD7Av1G011750	20905388	20908162	Growth-regulating factor	
5	AX-158591262	7A	<i>q.gwas.30</i>	TRITD7Av1G017240	30780092	30782694	Zinc finger CCCH domain protein TE?	DTH
6	AX-158591262	7A	<i>q.gwas.30</i>	TRITD7Av1G017550	31546753	31548244	NAC domain-containing protein, putative	
7	Ku_c24482_1132	7B	<i>q.gwas.35</i>	TRITD7Bv1G056500	157854628	157855010	Seed maturation protein LEA 4	
8	Ku_c24482_1132	7B	<i>q.gwas.35</i>	TRITD7Bv1G056720	158562417	158562866	Zinc finger family protein	
9	Ku_c24482_1132	7B	<i>q.gwas.35</i>	TRITD7Bv1G057630	162519251	162520668	FLOWERING LOCUS T/TERMINAL FLOWER 1-like protein	
10	Ku_c24482_1132	7B	<i>q.gwas.35</i>	TRITD7Bv1G058480	164897716	164902189	Phosphate transporter PHO1-like protein	
11	AX-109869840	6A	<i>q.gwas.25</i>	TRITD6Av1G220960	603258174	603260717	Ethylene receptor	
12	Tdurum_contig49186_437	7A	<i>q.gwas.30</i>	TRITD7Av1G011750	20905388	20908162	Growth-regulating factor	DTM
13	Tdurum_contig49186_437	7A	<i>q.gwas.30</i>	TRITD7Av1G017240	30780092	30782694	Zinc finger CCCH domain protein TE?	
14	Tdurum_contig49186_437	7A	<i>q.gwas.30</i>	TRITD7Av1G017550	31546753	31548244	NAC domain-containing protein, putative	
15	Kukri_c60966_261	6B	<i>q.gwas.28</i>	TRITD6Bv1G226900	693249887	693252855	Receptor protein kinase, Putative	GFP
16	AX-158602974	1B	<i>q.gwas.03</i>	TRITD1Bv1G189370	580660944	580663298	Calcineurin B-like protein	
17	AX-158602974	1B	<i>q.gwas.03</i>	TRITD1Bv1G189570	580988293	580988886	Receptor-like protein kinase	PHT
18	AX-158602974	1B	<i>q.gwas.03</i>	TRITD1Bv1G191400	585136543	585142125	Zinc finger protein	
19	BS00091519_51	5B	<i>q.gwas.24</i>	TRITD5Bv1G001780	5178666	5198798	Cytochrome P450-like protein	
20	Kukri_c17062_618	2A	<i>q.gwas.12</i>	TRITD2Av1G189490	526757053	526762842	Acyl-CoA N-acyltransferase	
21	Kukri_c17062_618	2A	<i>q.gwas.12</i>	TRITD2Av1G190600	529454657	529455487	Ring finger protein, Putative	SPL
22	Kukri_c3096_1411	2B	<i>q.gwas.13</i>	TRITD2Bv1G109560	316041878	316042468	E3 ubiquitin-protein ligase	
23	RAC875_c400_193	1B	<i>q.gwas.02</i>	TRITD1Bv1G000110	327500	328114	Tryptophan aminotransferase-related protein 2	
24	AX-89760660	1B	<i>q.gwas.07</i>	TRITD1Bv1G166820	518699664	518701642	Zinc finger CCCH domain-containing protein 4	
25	AX-89760660	1B	<i>q.gwas.07</i>	TRITD1Bv1G167110	519057378	519065351	UDP-GLUCOSE PYROPHOSPHORYLASE 1	SPP
26	Tdurum_contig25602_212	2B	<i>q.gwas.15</i>	TRITD2Bv1G184650	545766407	545768972	ethylene-responsive transcription factor	
27	AX-94631122	3A	<i>q.gwas.16</i>	TRITD3Av1G275580	723577014	723578195	E3 ubiquitin-protein ligase	
28	AX-158606713	1B	<i>q.gwas.08</i>	TRITD1Bv1G176830	544878459	544879181	Ethylene-responsive factor-like transcription factor	
29	AX-158606713	1B	<i>q.gwas.08</i>	TRITD1Bv1G177060	545904571	545908993	E3 ubiquitin-protein ligase	TKW
30	AX-158606713	1B	<i>q.gwas.08</i>	TRITD1Bv1G177540	547614469	547614996	Blue copper protein	
31	RFL_Contig3481_1669	1B	<i>q.gwas.02</i>	TRITD1Bv1G000110	327500	328114	Tryptophan aminotransferase-related protein 2	GYD
32	IAAV3365	5A	<i>q.gwas.21</i>	TRITD5Av1G205000	550462923	550477847	ABC transporter	

Chr, chromosome; GSS, Gene Sequence starts; GSE, Gene sequence ends; APQ, Associated putative QTL; DTH, Days to heading; DTM, Days to physiological maturity; GFP, Grain filling period; PHT, Plant height; SPL, Spike length; SPP, Number of spikelets per spike; TKW, Thousand kernel weight; GYD, Grain yield.

(Letta et al., 2014). This QTL *q.gwas.33* is also associated with genes *TRITD5BvG245710* (myb-like protein X), *TRITD5BvG246830* (KH domain containing protein), *TRITD5BvG247760* (NBS-LRR disease resistance protein-like protein) and *TRITD5BvG246270* (Glycosyltransferase).

The QTL region of *q.gwas.02* (RFL_Contig3481_1669) for GYD, identified in this study, overlaps with QTLs for several traits such as total root number and length (Maccaferri et al., 2016), grain protein content and concentration (Suprayogi et al., 2009), spikes per plant (Mengistu et al., 2016), heading date (Maccaferri et al., 2008), grain filling period (Soriano et al., 2016), semolina yellowness (Colasuonno et al., 2017), grain yield per spike and grain yield (Roncallo et al., 2018), grain protein content (Giraldo et al., 2016) and fusarium head blight resistance (Ghavami et al., 2011). The overlapping of QTLs for several important traits in this genomic region indicates its significance in future durum wheat breeding for grain yield and end-use quality traits. The genomic region corresponding to QTL *q.gwas.21* on chromosome 5A (MTA for IAAV3365 SNP and GYD) is a novel major QTL for grain yield, explaining the largest proportion of phenotypic variance ($r^2 = 44.95\%$) as compared to all other putative QTLs reported here. This is a highly significant result of this study, which needs to be validated through further research, including fine mapping to pinpoint the gene(s) responsible for this QTL. Interestingly, the genomic region of this QTL overlaps with previously identified QTL for number of kernels per spike (Kidane et al., 2017a), yellow rust resistance (Liu et al., 2017b), threshing time (Tzarfati et al., 2014), leaf rust resistances (Aoun et al., 2016), and total root number (Maccaferri et al., 2016). Therefore, this genomic region is a key target region for the improving of durum wheat, for grain yield and threats of wheat arising due to the impacts of climate change. The *TRITD5Av1G205000* (an ABC transporter) gene is one of the potential candidate genes behind this QTL (*q.gwas.21*). This is because previous research indicated that ABC transporter genes affect grain formation in wheat during heading and also modulate the ripening of the heads (Wanke and Üner Kolukisaoglu, 2010; Walter et al., 2015).

The present study identified several novel QTLs for grain yield-related traits, SPP and TKW. Additional MTAs that confirmed previously identified genomic regions were also detected for these traits. The three novel putative QTLs for SPP are *q.gwas.07* (721.4–725.8 Mbp) on chromosome 1B, *q.gwas.15* (541.4–551.4 Mbp) on chromosome 2B and *q.gwas.16* (721.4–725.8 Mbp) on chromosome 3A. For TKW, five novel putative QTLs, i.e., *q.gwas.18* (58.8–68.0 Mbp) on chromosome 3B, *q.gwas.19* (803.2–812.9 Mbp) on chromosome 3B, *q.gwas.22* (431.8–442.1 Mbp) on chromosome 5A, *q.gwas.26* (524–534 Mbp) on chromosome 6A, and *q.gwas.34* (681.9–692 Mbp) on chromosome 7A were identified. Fine mapping of these genomic regions is required to identify the genes responsible for these QTLs for SPP and TKW. However, for four of the five novel QTLs for SPP, we were able to identify potential candidate

genes, i.e., *TRITD1Bv1G000110* (Tryptophan aminotransferase-related protein 2), *TRITD1Bv1G167110* (UDP-Glucose Pyrophosphorylase 1), *TRITD2Bv1G184650* (ethylene-responsive transcription factor), and *TRITD3Av1G275580* (E3 ubiquitin-protein ligase). *TRITD2Bv1G184650* has been reported to regulate the initiation and development of spikelets in wheat, particularly when the temperature is low (Yu et al., 2021).

Several putative QTLs for SPP are identified here, i.e., *q.gwas.02* (0.3–5.8 Mbp), *q.gwas.14* (94–104.6 Mbp), *q.gwas.20* (721.4–725.8 Mbp), and *q.gwas.30* (2–35 Mbp) were found co-localized with previously reported QTLs for these traits on chromosomes 1B, 2B, 4A, and 7A, respectively (Golabadi et al., 2011; Mengistu et al., 2016; Kidane et al., 2017a; Roncallo et al., 2017; Mangini et al., 2018; Soriano et al., 2018; Li et al., 2019; Rahimi et al., 2019; Alipour et al., 2021). Similarly, putative QTLs for TKW were co-localized with QTLs previously identified, i.e., *q.gwas.08* (MTA for AX-158606713) with a QTL identified based on Ethiopian durum wheat germplasm (Mengistu et al., 2016), and *q.gwas.30* with a QTL identified by Golabadi et al. (2011) based on F3 and F4 populations of durum wheat in Iran, and by Mangini et al. (2018) from a collection of tetraploid wheat grown in Southern Italy. The genomic region regarded as QTL *q.gwas.30* in this study was associated with four traits (DTH, DTM, SPP, TKW) (Table 5). This suggests that either the same gene with pleiotropic effects is involved in regulating these traits, or different genes in this genomic region regulate their corresponding traits or a combination of both. Thus, further research is required to identify common SNP markers representing the four traits in this genomic region and subsequent use in marker-assisted selection for improving the crop. Several genes encoding growth-regulating factor, seed maturation protein, phosphate transporter, phototropic-responsive NPH3 protein G, disease resistance protein RPM1, phosphate-responsive 1 family protein, E3-ubiquitin-protein ligase SINA-like 10, potassium transporter, and chloroplast envelope membrane protein, are among the likely candidates for the QTL *q.gwas.30* (Supplementary Table 7).

The marker-trait association analysis conducted via GWAS discovered novel and previously identified genomic regions (putative QTLs) associated with DTH. Of these, *q.gwas.01* (522.7–528.6 Mbp of chromosome 1B), *q.gwas.09* (753.4–757.6 Mbp of chromosome 2A), and *q.gwas.30* (2–35 Mbp of chromosome 3B) were previously reported for this trait (Golabadi et al., 2011; Roncallo et al., 2017; Mangini et al., 2018). These QTLs are significant at two or more test locations and hence can be considered stable MTAs across environments. The present findings also confirmed the results reported in previous studies for DTH on chromosomes 1B, 2A, and 3B (Kidane et al., 2017b; Ogbonnaya et al., 2017; Li et al., 2019; Rahimi et al., 2019; Wang et al., 2019). Hence, there is solid evidence of genes regulating DTH in these genomic regions. One of the novel putative QTLs for DTH is *q.gwas.35*, covering a

151.9–165.3 Mbp region on chromosome 7B. This genomic region contains the *TRITD7Bv1G057630* gene that encodes Flowering Locus T/Terminal Flower 1-like protein. Previous research on wheat, soybean, and Arabidopsis indicates that this gene is located in the region flanking *FT-D1*, a major gene regulating flowering in wheat, soybean, and Arabidopsis (Sun et al., 2019; Isham et al., 2021). Hence, if breeders aim to improve durum wheat for DTH, it is advisable to consider the QTL regions of *q.gwas.35* to get information related to DTH. Furthermore, as previously shown (Wu et al., 2008; Zhao et al., 2020), the *q.gwas.01* QTL region contains a potential candidate gene *TRITD1Bv1G168480* (WRKY). This gene involved in regulating leaf senescence. It is also known to have major roles at various stages of wheat development affecting productivity and product quality and was predicted to interact with DTH. Similarly, the *TRITD7Av1G017240* (zinc finger CCCH-type transcription factor) gene, which promotes wheat flowering, was also identified in this study. Hence, it would be worthwhile to conduct further research on this genomic region to identify the gene involved in *q.gwas.01* and to understand the relationship between leaf senescence and DTH in durum wheat.

A previously known genomic region and three novel genomic regions (putative QTLs) associated with DTM were found on chromosomes 2A, 6A, and 7A. These QTLs explained 2–33% of the variation in DTM. The QTL designated as *q.gwas.25* (601.5–615.3 Mbp of chromosome 6A) was reported in a previous study on wheat (Mangini et al., 2018). The novel putative QTLs are located on chromosomes 2A (67.5–70.6 Mbp; *q.gwas.10*) and 7A (2.1–35.1 Mbp; *q.gwas.30*, and 35.5–68.8 Mbp; *q.gwas.31*). The *TRITD6Av1G220960* (Ethylene receptor) gene, which is located within the genomic region of *q.gwas.25*, is a potential candidate gene for *q.gwas.25*. Previous research has suggested that ethylene receptors are most likely related to the duration of seed development and maturation; i.e., the duration embryo development (Hays et al., 2007). However, in previous findings in maize, grain yield increments were observed through ethylene signal reduction (Shi et al., 2015). Similar to the previous study (Han et al., 2021), *TRITD7Av1G011750* (growth-regulating factor), *TRITD6Av1G017240* (Zinc finger CCCH domain protein TE), and *TRITD7Av1G017550* (NAC domain-containing protein) genes have been found in the genomic regions of QTLs for DTH identified in this study, and are potential candidate genes for the corresponding QTLs. Studies have shown that these genes are mainly involved in regulating growth, development, biotic and abiotic stress adaptation in wheat, rice, and other crop plants and may also determine the variation in phenological traits in wheat and rice.

MTA analysis for PHT identified four putative QTLs, *q.gwas.03* (580.6–585.4 Mbp of chromosomes 1B), *q.gwas.04* (669.5–674.7 Mbp of chromosomes 1B) and *q.gwas.24* (628.2–630.9 Mbp of chromosomes 5B), similar to the previous reports (Zhang et al., 2012; Roncallo et al., 2017). This study also found novel MTAs and putative QTLs on chromosome 1B (*q.gwas.05*;

628.2 – 630.9 Mbp), 5A (*q.gwas.22*; 431.8–442 Mbp), 6B (*q.gwas.29*; 119.6–121.5 Mbp), and 7B (*q.gwas.37*; 686.4–697.9 Mbp). These results show that all significant MTAs for PHT were on the B genome except one MTA on 5A. This may serve as an indicator of potential hotspot regions for genes associated with PHT. The *TRITD1Bv1G191400* (Zink finger protein) could be a candidate gene underlying the *q.gwas.03* QTL for PHT. Previous research revealed that this gene is significantly associated with improved salt tolerance and regulates stress resistance in wheat, Arabidopsis, and other plants (Ciftci-Yilmaz and Mittler, 2008; Ma et al., 2016). Other potential candidate genes for the PHT QTLs include *TRITD1Bv1G189370* (encoding Calcineurin B-like protein), *TRITD1Bv1G189570* (encoding Receptor-like protein kinase), and *TRITD5Bv1G0001780* (Cytochrome P450-like protein).

The present study confirmed previously reported QTL for SPL on chromosome 1B (*q.gwas.06*; 314.3–318.8 Mbp), which was previously reported by Giraldo et al. (2016). Likewise, putative QTL *q.gwas.11* on chromosome 2A (489.5–492.5 Mbp), QTL *q.gwas.12* on chromosome 2A (522.4–534.5 Mbp), QTL *q.gwas.13* on chromosome 2B (185.5–195.6 Mbp), and QTL *q.gwas.23* on chromosome 5A (526.3–534.1 Mbp) were found for this trait. These are likely to be novel QTLs since the corresponding genomic regions are not associated with SPL in previous studies. The potential candidate genes for the SPL QTLs include *TRITD2Av1G189490* (encoding Acyl-CoA N-acyltransferase), *TRITD2Av1G190600* (encoding Ring finger protein), and *TRITD2Bv1G109560* (encoding E3 ubiquitin-protein ligase). A report from a previous study revealed that E3 ubiquitin proteins have a potential role in modulating crop productivity by influencing growth, development, and important agronomic traits (Varshney and Majee, 2022).

5 Conclusions

In the present study, we evaluated the diverse germplasm of Ethiopian durum wheat using multi-environment trials (MET) data that are genotyped with the Illumina Infinium 25k wheat SNP array to unravel genomic regions associated with its phenological and plant architecture traits as well as grain yield and yield related traits using GWAS. The GWAS identified 44 significant MTAs, including 26 novel genomic regions. The combined analysis of variance revealed significant effects of genotype, environment, and genotype-by-environment interaction on the target traits. The study also confirmed several previously reported QTLs. The identification of a large number of novel QTLs in this study indicates the presence of novel alleles of the genes underlying these QTLs, which probably confirms the distinctness of the Ethiopian durum wheat gene pool from other durum wheat gene pools. The major significant QTLs, such as *q.gwas.21* (for SNP IAAV3365, stable across location) that explained 44.95% of the variation in grain yield, *q.gwas.10* (for SNP Kukri_rep_c73477_888)

that explained 32.9% of the variation in days to maturity and *q.gwas.28* (for SNP Kukri_c60966_261) that explained 30.7% of the variation in the grain-filling period are the key findings of this study. Additional research is needed to validate these key findings, including fine mapping to determine the underlying genes and their subsequent functional analysis. The addition of SNP markers associated with the target traits of this study is highly beneficial for genomic-led breeding of durum wheat.

The results could empower the sustainability of durum breeding by unlocking genomic regions governing complex plant characteristics. Most importantly, the results obtained in the present study could contribute a major role in understanding the durum wheat genome and improving genetic resources for breeding this crop, which in turn, supports global food security. The newly identified genes will also advance the understanding of genomic regions associated with essential characteristics used in durum wheat breeding. The identified novel variants suggest a potential use of Ethiopian durum wheat in durum wheat marker-assisted breeding. The study also provided new insight into the genetic architecture of grain yield and related traits. It indicated the potential of the diverse Ethiopian durum wheat gene pool for future improvement programs. Hence, the identified MTAs and candidate genes could be used to understand the genetic basis of genomic regions of important traits and to accelerate the development of new cultivars with high grain yield and agronomically essential traits *via* precision breeding. In addition, the identified MTAs could be used in marker-assisted breeding, fine mapping, and cloning of the underlying genomic regions and putative QTLs in durum wheat germplasm.

Data availability statement

The genotypic data presented in this study were generated using a commercially available Illumina Infinium 25k wheat SNP array whose details can be found at <https://www.traitgenetics.com/index.php/service-products>. The genotypic data of the 420 genotypes studied can be obtained upon request.

Author contributions

Conceptualization: BM, KT, RO, MG, EJ, TH, Methodology: BM, MG, KT, RO, EJ, TH, CH, Data curation: BM, Formal analysis: BM, Visualization: BM, MG, Investigation: BM, MG, RO, EJ, Resources: KT, RO, EJ, MG, TH, Funding acquisition: KT, RO, EJ, MG, TH, Project administration: KT, RO, MG, EJ, TH, Supervision: KT, RO, EJ, MG, TH, CH, FH, Writing original draft: BM, Writing-review, and editing: BM, RO, EJ, MG, KT, TH, CH, and FH. All authors have read and approved the final version of the manuscript.

Funding

This work was funded by the Swedish International Development Cooperation Agency (Sida) grant awarded to Addis Ababa University and the Swedish University of Agricultural Sciences for a bilateral capacity-building program in biotechnology. The funding information is available on "<https://sida.aau.edu.et/index.php/biotechnology-phdprogram/>; accessed on May 21, 2022". The funders played no role in the design of the study, data collection, analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

The authors would like to thank the Institute of Biotechnology, Addis Ababa University (AAU) for the technical support received during the fieldwork and the Swedish University of Agricultural Science for different facility support during seedling development in the greenhouse. The authors are grateful to Sinana Agricultural Research Center (SARC), Kulumsa Agricultural Research Center (KARC), Debrezeit Agricultural Research Center (DZARC), and Wolmera Agricultural Office for providing their experimental sites and equipment for the field trials. The unreserved support from the technical staff of SARC, KARC, and DZARC during the field trial is highly appreciated. We would also like to thank Ethiopian Biodiversity Institute, Wollo University, SARC, and DZARC for providing durum wheat germplasm.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1009244/full#supplementary-material>

SUPPLEMENTARY TABLE 2

Description of mean monthly weather information of test sites during crop growing season.

SUPPLEMENTARY TABLE 5

Computed physical distance at which mean linkage disequilibrium decay drops below at cut off ($r^2 = 0.2$) value on pairwise comparison of SNP markers.

SUPPLEMENTARY TABLE 6

GWAS complete output for phenological, plant architecture, grain yield, and yield-related traits across five test sites and combined data. "Trait" indicate measured phenotypic data including; DTH, Days to heading; DTM, Days to physiological maturity; NET, Number of effective tillers per plant; GFP, Grain filling period; PHT, Plant height; SPL, Spike length; SPP, Number of spikelets per spike; TKW, Thousand kernel weight; and GYD, Grain yield. Loc, test sites, AK, CD, HO, KU, and SN are Akaki, Chefe Donsa, Holeta, Kulumsa, and Sinana test sites, respectively. CO; reports combined environments data; CHR, chromosome; SNP, Single nucleotide polymorphisms; POS, Physical position; MAF, minor allele frequency; P.value, reports the significance of the nominal tests and PVE, reports phenotypic variance explained. This [Supplementary Table](#) is also presented in Excel format.

SUPPLEMENTARY TABLE 7

High-confidence candidate genes were identified through marker-trait association analysis to associate with phenological, plant architecture, grain yield, and yield-related traits collected at five test sites. "Trait" indicates measured phenotypic data including; DTH, Days to heading; DTM, Days to physiological maturity; NET, Number of effective tillers per plant; GFP, Grain filling period; PHT, Plant height; SPL, Spike length; SPP, Number of spikelets per spike; TKW, Thousand kernel weight; and GYD, Grain yield. MTAs, Marker trait associations; SNPs, Single nucleotide polymorphisms. This [Supplementary Table](#) is also presented in Excel format.

SUPPLEMENTARY FIGURE 1

Genome-wide LD decay plot over total physical distance based on 10,045 SNP markers. The yellow curve represents the model that fits LD decay. The solid red line represents the arbitrary threshold for no LD used ($r^2 = 0.3$). The light green line indicates the intersection between the critical and the map distance to determine QTL confidence intervals.

SUPPLEMENTARY FIGURE 2

The circular Manhattan and Q-Q plots of GWAS results for DTH on panels (A, B), and DTM on panels (C, D) were produced using each test site and combined data across test sites. The circular Manhattan plots represent the relative positions of the SNP markers on each chromosome in a circular manner. To view the significant MTAs results, move from outside to the center of each circle, rotating (rounding) through each circle starting from FarmCPU.CO.DTH, followed by FarmCPU.SN.DTH, FarmCPU.KU.DTH, FarmCPU.HO.DTH, FarmCPU.CD.DTH, and FarmCPU.AK.DTH to the center of the circle in sequential order and follow a similar approach for DTM. The name of test sites as stated here is presented in the center of circle. For the QQ-plots, Y-axis represents observed $-\log_{10}$ (p-value), and the X-axis represents expected $-\log_{10}$ (p-value) under the assumption that the p-values follow a normal distribution. The dotted lines indicate the 95% confidence interval assuming the null hypothesis of no association between the SNP and trait. DTH refers to days to heading, and DTM refers to days to physiological maturity. CO; Combined data across five environments, SN: Sinana site, KU: Kulumsa site, HO: Holeta site, CD: Chefe Donsa site, and AK: Akaki sites. The Circular Manhattan and Q-Q plots of GWAS results for GYD on panel (E, F), TKW on panel (G, H), and SPP on panel (I, J) were plotted using data from each test site and combined data from all test sites, respectively. For the circular manhattan plot, follow a similar approach for these traits as in [Figure S2A](#) for all test sites and traits. The assumption of QQ-plots in also applies here. GYD, Grain yield, TKW, Thousand-kernel weight, and SPP, number of spikelets per spike. CO; Combined data across five environments, SN: Sinana site, KU: Kulumsa site, HO: Holeta site, CD: Chefe Donsa site, and AK: Akaki sites. The Circular Manhattan and Q-Q plots of GWAS results for PHT on panel (K, L), NET on panel (M, N), and SPL on panel (O, P) were plotted using data from each test site and combined data from all test sites, respectively. For the circular manhattan plot, follow a similar approach for these traits as in [Figure S 2A](#) for all test sites and traits. For the QQ-plots, apply the assumption in a. SPL, Spike length; NET, Number of effective tillers per plant; and PHT, Plant height, and SPP. CO; Combined data across five environments, SN, Sinana site; KU, Kulumsa site; HO, Holeta site; CD, Chefe Donsa site; and AK, Akaki sites.

References

- Abu-Zaitoun, S. Y., Chandrasekhar, K., Assili, S., Shtaya, M. J., Jamous, R. M., Mallah, O. B., et al. (2018). Unlocking the genetic diversity within a middle-east panel of durum wheat landraces for adaptation to semi-arid climate. *Agronomy* 8(10), 233. doi: 10.3390/agronomy8100233
- Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., et al. (2018). Linking the international wheat genome sequencing consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.* 19(1), 111. doi: 10.1186/s13059-018-1491-4
- Alemu, A., Brazauskas, G., Gaikpa, D. S., Henriksson, T., Islamov, B., Jørgensen, L. N., et al. (2021a). Genome-wide association analysis and genomic prediction for adult-plant resistance to *Septoria tritici* blotch and powdery mildew in winter wheat. *Front. Genet.* doi: 10.3389/fgene.2021.661742
- Alemu, A., Feyissa, T., Letta, T., and Abeyo, B. (2020a). Genetic diversity and population structure analysis based on the high density SNP markers in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *BMC Genet.* 21(1), 18. doi: 10.1186/s12863-020-0825-x
- Alemu, A., Feyissa, T., Maccaferri, M., Sciara, G., Tuberosa, R., Ammar, K., et al. (2021b). Genome-wide association analysis unveils novel QTLs for seminal root system architecture traits in Ethiopian durum wheat. *BMC Genomics.* 22, 20. doi: 10.1186/s12864-020-07320-4
- Alemu, A., Feyissa, T., Tuberosa, R., Maccaferri, M., Sciara, G., Letta, T., et al. (2020b). ScienceDirect genome-wide association mapping for grain shape and color traits in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *Crop J.* 8(5), 757–768. doi: 10.1016/j.cj.2020.01.001
- Alemu, S. K., Huluka, A. B., Tesfaye, K., Haileselassie, T., and Uauy, C. (2021). Genome-wide association mapping identifies yellow rust resistance loci in Ethiopian durum wheat germplasm. *PLoS One.* 16(5), e0243675. doi: 10.1371/journal.pone.0243675
- Alipour, H., Abdi, H., Rahimi, Y., and Bihamta, M. R. (2021). Dissection of the genetic basis of genotype-by-environment interactions for grain yield and main agronomic traits in Iranian bread wheat land-races and cultivars. *Sci Rep.* 11(1):17742. doi: 10.1038/s41598-021-96576-1.
- Al-Khayri, J. M., Jain, S. M., and Johnson, D. V. (2016). Advances in plant breeding strategies: Breeding, biotechnology and molecular tools. Switzerland, Springer International Publishing, 656 p. doi: 10.1007/978-3-319-22521-0
- Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueno, J., Crossa, J., Vargas, M., et al. (2020). META-r: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8(5), 745–756. doi: 10.1016/j.cj.2020.03.010
- Anuarbek, S., Abugaliyeva, S., Pecchioni, N., Laido, G., Maccaferri, M., Tuberosa, R., et al. (2020). Quantitative trait loci for agronomic traits in tetraploid wheat for

- enhancing grain yield in Kazakhstan environments. *PLoS One*. 5(6), e0234863. doi: 10.1371/journal.pone.0234863
- Aoun, M., Breiland, M., Kathryn Turner, M., Loladze, A., Chao, S., Xu, S. S., et al. (2016). Genome-wide association mapping of leaf rust response in a durum wheat worldwide germplasm collection. *Plant Genome*. 9(3). doi: 10.3835/plantgenome2016.01.0008
- Aoun, M., Rouse, M. N., Kolmer, J. A., Kumar, A., and Elias, E. M. (2021). Genome-wide association studies reveal all-stage rust resistance loci in elite durum wheat genotypes. *Front. Plant Sci.* 12, 640739. doi: 10.3389/fpls.2021.640739
- Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Software*. doi: 10.18637/jss.v067.i01
- Bellucci, A., Tondelli, A., Fangel, J. U., Torp, A. M., Xu, X., Willats, W. G. T., et al. (2017). Genome-wide association mapping in winter barley for grain yield and culm cell wall polymer content using the high-throughput CoMPP technique. *PLoS One*. 12(3), e0173313. doi: 10.1371/journal.pone.0173313
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*. 57(1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bhatta, M., Morgounov, A., Belamkar, V., and Baenziger, P. S. (2018). Genome-wide association study reveals novel genomic regions for grain yield and yield-related traits in drought-stressed synthetic hexaploid wheat. *Int. J. Mol. Sci.* 19(10), 3237. doi: 10.3390/ijms19103011
- Borrego-Benjumea, A., Carter, A., Zhu, M., Tucker, J. R., Zhou, M., and Badaea, A. (2021). Genome-wide association study of waterlogging tolerance in barley (*Hordeum vulgare* L.) under controlled field conditions. *Front. Plant Sci.* 12, 711654. doi: 10.3389/fpls.2021.711654
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Bresaghella, F., and Sorrells, M. E. (2006a). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46(3), 1323–1330. doi: 10.2135/cropsci2005.09-0305
- Bresaghella, F., and Sorrells, M. E. (2006b). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*. 172(2), 1165–1177. doi: 10.1534/genetics.105.044586
- Canè, M. A., Maccaferri, M., Nazemi, G., Salvi, S., Francia, R., Colalongo, C., et al. (2014). Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Mol. Breed.* 34(4), 1629–1645. doi: 10.1007/s11032-014-0177-1
- Ceglar, A., Toreti, A., Zampieri, M., and Royo, C. (2021). Global loss of climatically suitable areas for durum wheat growth in the future. *Environ. Res. Lett.* 16, 104049. doi: 10.1088/1748-9326/ac2d68
- Chen, J., Zhang, F., Zhao, C., Lv, G., Sun, C., Pan, Y., et al. (2019). Genome-wide association study of six quality traits reveals the association of the TaRPP13L1 gene with flour colour in Chinese bread wheat. *Plant Biotechnol. J.* 17(11), 2106–2122. doi: 10.1111/pbi.13126
- Ciftci-Yilmaz, S., and Mittler, R. (2008). The zinc finger network of plants. *Cell. Mol. Life Sci.* 65, 1150–1160. doi: 10.1007/s00018-007-7473-4
- Colasuonno, P., Lozito, M. L., Marcotuli, I., Nigro, D., Giancaspro, A., Mangini, G., et al. (2017). The carotenoid biosynthetic and catabolic genes in wheat and their association with yellow pigments. *BMC Genomics*. 18(1), 122. doi: 10.1186/s12864-016-3395-6
- Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* 363(1491), 557–572. doi: 10.1098/rstb.2007.2170
- Condorell, G. E., Newcomb, M., Grolí, E. L., Maccaferri, M., Forestan, C., Babaeian, E., et al. (2022). Genome wide association study uncovers the QTLome for osmotic adjustment and related drought adaptive traits in durum wheat. *Genes (Basel)*. 13(2), 293. doi: 10.3390/genes13020293
- Corbeil, R. R., and Searle, S. R. (1976). Restricted maximum likelihood (reml) estimation of variance components in the mixed model. *Technometrics*. 18, 31–38. doi: 10.1080/00401706.1976.10489397
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14(8), 2611–20. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes*. 7(4), 574–578. doi: 10.1111/j.1471-8286.2007.01758.x
- Fayaz, F., Aghaee Sarbarzeh, M., Talebi, R., and Azadi, A. (2019). Genetic diversity and molecular characterization of Iranian durum wheat landraces (*Triticum turgidum* durum (Desf.) husn.) using DArT markers. *Biochem. Genet.* 57(1), 98–116. doi: 10.1007/s10528-018-9877-2
- Fiedler, J. D., Salsman, E., Liu, Y., Michalak de Jiménez, M., Hegstad, J. B., Chen, B., et al. (2017). Genome-wide association and prediction of grain and semolina quality traits in durum wheat breeding populations. *Plant Genome*. 10(3). doi: 10.3835/plantgenome2017.05.0038
- Gao, L., Meng, C., Yi, T., Xu, K., Cao, H., Zhang, S., et al. (2021). Genome-wide association study reveals the genetic basis of yield- and quality-related traits in wheat. *BMC Plant Biol.* 21(1), 144. doi: 10.1186/s12870-021-02925-7
- Garcia, M., Eckermann, P., Haeefe, S., Satija, S., Sznajder, B., Timmins, A., et al. (2019). Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS One*. 14(2), e0211730. doi: 10.1371/journal.pone.0211730
- Geleta, M., and Ortiz, R. (2016). Molecular and Genomic Tools Provide Insights on Crop Domestication and Evolution. In: L. S. Donald, Editor(s). *Advances in Agronomy*, Academic Press 135, 185–223. doi: 10.1016/bs.agron.2015.09.005.
- Ghavami, F., Elias, E. M., Mamidi, S., Mergoum, M., Kianian, S. F., Ansari, O., et al. (2011). Mixed model association mapping for fusarium head blight resistance in tunisian-derived durum wheat populations. *G3 Genes Genomes Genet.* 1(3), 209–218. doi: 10.1534/g3.111.000489
- Giraldo, P., Royo, C., González, M., Carrillo, J. M., and Ruiz, M. (2016). Genetic diversity and association mapping for agromorphological and grain quality traits of a structured collection of durum wheat landraces including subsp. durum, turgidum and diccocon. *PLoS One*. 11(11), e0166577. doi: 10.1371/journal.pone.0166577
- Golabadi, M., Arzani, A., Mirmohammadi Maibody, S. A. M., Tabatabaei, B. E. S., and Mohammadi, S. A. (2011). Identification of microsatellite markers linked with yield components under drought stress at terminal growth stages in durum wheat. *Euphytica*. 177, 207–221. doi: 10.1007/s10681-010-0242-8
- Gonçalves-Vidigal, M. C., Mora, F., Bignotto, T. S., Munhoz, R. E. F., and De Souza, L. D. (2008). Heritability of quantitative traits in segregating common bean families using a Bayesian approach. *Euphytica* 164, 551–560. doi: 10.1007/s10681-008-9758-6
- Han, G., Qiao, Z., Li, Y., Wang, C., and Wang, B. (2021). The roles of ccch zinc-finger proteins in plant abiotic stress tolerance. *Int. J. Mol. Sci.* 22(15), 8327. doi: 10.3390/ijms22158327
- Hays, D. B., Do, J. H., Mason, R. E., Morgan, G., and Finlayson, S. A. (2007). Heat stress induced ethylene production in developing wheat grains induces kernel abortion and increased maturation in a susceptible cultivar. *Plant Sci.* 172, 1113–1123. doi: 10.1016/J.PLANTSCI.2007.03.004
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695.
- IBPGR (1985). *Descriptors wheat*, Vol. 12.
- Isham, K., Wang, R., Zhao, W., Wheeler, J., Klassen, N., Akhunov, E., et al. (2021). QTL mapping for grain yield and three yield components in a population derived from two high-yielding spring wheat cultivars. *Theor. Appl. Genet.* 134(7), 2079–2095. doi: 10.1007/s00122-021-03806-1
- Johansson, E., Branlard, G., Cuniberti, M., Flagella, Z., Hüsken, A., Nurit, E., et al. (2020a). “Genotypic and environmental effects on wheat technological and nutritional quality,” in *Wheat Quality For Improving Processing And Human Health*. ed. T. M. Igrejas, C. Ikeda and G. Guzmán (Springer Cham, Springer Nature Switzerland AG), 171–204. doi: 10.1007/978-3-030-34163-3_8
- Johansson, E., Henriksson, T., Prieto-Linde, M. L., Andersson, S., Ashraf, R., and Rahmatov, M. (2020b). Diverse wheat-alien introgression lines as a basis for durable resistance and quality characteristics in bread wheat. *Front. Plant Sci.* 11, 1067. doi: 10.3389/fpls.2020.01067
- Johansson, E., Prieto-Linde, M. L., and Larsson, H. (2021). Locally adapted and organically grown landrace and ancient spring cereals—a unique source of minerals in the human diet. *Foods*. 10(2), 393. doi: 10.3390/foods10020393
- Kabbaj, H., Sall, A. T., Al-Abdallat, A., Geleta, M., Amri, A., Filali-Maltouf, A., et al. (2017). Genetic diversity within a global panel of durum wheat (*Triticum durum*) landraces and modern germplasm reveals the history of alleles exchange. *Front. Plant Sci.* 8, 1277. doi: 10.3389/fpls.2017.01277
- Kankwatsa, P., Singh, D., Thomson, P. C., Babiker, E. M., Bonman, J. M., Newcomb, M., et al. (2017). Characterization and genome-wide association mapping of resistance to leaf rust, stem rust and stripe rust in a geographically diverse collection of spring wheat landraces. *Mol. Breed.* 37(3), 113. doi: 10.1007/s11032-017-0707-8
- Kidane, Y. G., Gesesse, C. A., Hailemariam, B. N., Desta, E. A., Mengistu, D. K., Fadda, C., et al. (2019). A large nested association mapping population for breeding and quantitative trait locus mapping in Ethiopian durum wheat. *Plant Biotechnol. J.* 17(7), 1380–1393. doi: 10.1111/pbi.13062
- Kidane, Y. G., Hailemariam, B. N., Mengistu, D. K., Fadda, C., Pè, M. E., and Dell’Acqua, M. (2017a). Genome-wide association study of septoria tritici blotch

- resistance in Ethiopian durum wheat landraces. *Front. Plant Sci.* 8, 1586. doi: 10.3389/fpls.2017.01586
- Kidane, Y. G., Mancini, C., Mengistu, D. K., Frascaroli, E., Fadda, C., Pè, M. E., et al. (2017b). Genome wide association study to identify the genetic base of smallholder farmer preferences of durum wheat traits. *Front. Plant Sci.* 8, 1230. doi: 10.3389/fpls.2017.01230
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15(3), 1179–1191. doi: 10.1111/1755-0998.12387
- Kumar, D., Kumar, A., Chhokar, V., Gangwar, O. P., Bhardwaj, S. C., Sivasamy, M., et al. (2020). Genome-wide association studies in diverse spring wheat panel for stripe, stem, and leaf rust resistance. *Front. Plant Sci.* 11, 748. doi: 10.3389/fpls.2020.00748
- Letta, T., Olivera, P., Maccaferri, M., Jin, Y., Ammar, K., Badebo, A., et al. (2014). Association mapping reveals novel stem rust resistance loci in durum wheat at the seedling stage. *Plant Genome*. 7, 1. doi: 10.3835/plantgenome2013.08.0026
- Levene, H. (1960). Robust tests for equality of variances. *Contrib. to Probab. Stat. Essays*.
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*. 28(18), 2397–2397. doi: 10.1093/bioinformatics/bts444
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12(2), e1005767. doi: 10.1371/journal.pgen.1005767
- Liu, W., Maccaferri, M., Bulli, P., Rynearson, S., Tuberosa, R., Chen, X., et al. (2017a). Genome-wide association mapping for seedling and field resistance to puccinia striiformis f. sp. tritici in elite durum wheat. *Theor. Appl. Genet.* 130(4), 649–667. doi: 10.1007/s00122-016-2841-9
- Liu, W., Maccaferri, M., Chen, X., Laghetti, G., Pignone, D., Pumphrey, M., et al. (2017b). Genome-wide association mapping reveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum* ssp. dicoccum). *Theor. Appl. Genet.* 130, 2249–2270. doi: 10.1007/s00122-017-2957-6
- Liu, W., Maccaferri, M., Rynearson, S., Letta, T., Zegeye, H., Tuberosa, R., et al. (2017c). Novel sources of stripe rust resistance identified by genome-wide association mapping in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *Front. Plant Sci.* 8, 774. doi: 10.3389/fpls.2017.00774
- Liu, J., Rasheed, A., He, Z., Imtiaz, M., Arif, A., Mahmood, T., et al. (2019). Genome-wide variation patterns between landraces and cultivars uncover divergent selection during modern wheat breeding. *Theor. Appl. Genet.* 132(9), 2509–2523. doi: 10.1007/s00122-019-03367-4
- Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., et al. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* 19(1), 168. doi: 10.1186/s12870-019-1781-3
- Maccaferri, M., Cane, M. A., Sanguineti, M. C., Salvi, S., Colalongo, M. C., Massi, A., et al. (2014). A consensus framework map of durum wheat (*Triticum durum* desf.) suitable for linkage disequilibrium analysis and genome-wide association mapping. *BMC Genomics*. 15(1), 873. doi: 10.1186/1471-2164-15-873
- Maccaferri, M., El-Feki, W., Nazemi, G., Salvi, S., Canè, M. A., Colalongo, M. C., et al. (2016). Prioritizing quantitative trait loci for root system architecture in tetraploid wheat. *J. Exp. Bot.* 67(4), 1161–1178. doi: 10.1093/jxb/erw039
- Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., et al. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* 51, 885–895. doi: 10.1038/s41588-019-0381-3
- Maccaferri, M., Sanguineti, M. C., Corneti, S., Ortega, J. L. A., Salem, M.B., Bort, J., et al. (2008). Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* desf.) across a wide range of water availability. *Genetics*. 178(1), 489–511. doi: 10.1534/genetics.107.077297
- Maccaferri, M., Sanguineti, M. C., Mantovani, P., Demontis, A., Massi, A., Ammar, K., et al. (2010). Association mapping of leaf rust response in durum wheat. *Mol. Breed.* 26, 189–228. doi: 10.1007/s11032-009-9353-0
- Ma, X., Liang, W., Gu, P., and Huang, Z. (2016). Salt tolerance function of the novel C2H2-type zinc finger protein TaZNF in wheat. *Plant Physiol. Biochem.* 06, 129–140. doi: 10.1016/j.plaphy.2016.04.033
- Mangini, G., Gadaleta, A., Colasuonno, P., Marcotuli, I., Signorile, A. M., Simeone, R., et al. (2018). Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PLoS One*. 13(1), e0190162. doi: 10.1371/journal.pone.0190162
- Mathew, I., Shimelis, H., Shayanowako, A. I. T., Laing, M., and Chaplot, V. (2019). Genome-wide association study of drought tolerance and biomass allocation in wheat. *PLoS One*. 14(12), e0225383. doi: 10.1371/journal.pone.0225383
- Mazzucotelli, E., Sciara, G., Mastrangelo, A. M., Desiderio, F., Xu, S. S., Farris, J., et al. (2020). The global durum wheat panel (GDP): An international platform to identify and exchange beneficial alleles. *Front. Plant Sci.* 11, 569905. doi: 10.3389/fpls.2020.569905
- Mekonnen, T., Sneller, C. H., Haileselassie, T., Ziyomo, C., Abeyo, B. G., Goodwin, S. B., et al. (2021). Genome-wide association study reveals novel genetic loci for quantitative resistance to septoria tritici blotch in wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 12, 671323. doi: 10.3389/fpls.2021.671323
- Mengistu, K., Catellani, M., Frascaroli, E., Fadda, C., Pè, M. E., and Dell'Acqua, M. (2016). High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnol. J.* 14(9), 1800–1812. doi: 10.1111/pbi.12538
- Mengistu, D. K., Kiros, A. Y., and Pè, M. E. (2015). Phenotypic diversity in Ethiopian durum wheat (*Triticum turgidum* var. durum) landraces. *Crop J.* 3(3), 190–199. doi: 10.1016/j.cj.2015.04.003
- Mérida-García, R., Bentley, A. R., Gálvez, S., Dorado, G., Solís, I., Ammar, K., et al. (2020). Mapping agronomic and quality traits in elite durum wheat lines under differing water regimes. *Agronomy*. 10(1), 144. doi: 10.3390/agronomy10010144
- Mérida-García, R., Liu, G., He, S., Gonzalez-Dugo, V., Dorado, G., Gálvez, S., et al. (2019). Genetic dissection of agronomic and quality traits based on association mapping and genomic selection approaches in durum wheat grown in southern Spain. *PLoS One*. 14(2), e0211718. doi: 10.1371/journal.pone.0211718
- Mohammadi, R., Armion, M., Zadhanan, E., Ahmadi, M. M., and Amri, A. (2018). The use of AMMI model for interpreting genotype × environment interaction in durum wheat. *Exp. Agric.* 54(5), 670–683. doi: 10.1017/S0014479717000308
- Ogbonnaya, F. C., Rasheed, A., Okechukwu, E. C., Jighly, A., Makdis, F., Wuletaw, T., et al. (2017). Genome-wide association study for agronomic and physiological traits in spring wheat evaluated in a range of heat prone environments. *Theor. Appl. Genet.* 130, 1819–1835. doi: 10.1007/s00122-017-2927-z
- Ozkan, H., Brandolini, A., Schäfer-Pregl, R., and Salamini, F. (2002). AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. *Mol. Biol. Evol.* 19 (10), 1797–801. doi: 10.1093/oxfordjournals.molbev.a004002
- Patterson, H. D., and Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika*. 63(1), 83–92. doi: 10.1093/biomet/63.1.83
- Peng, J., Ronin, Y., Fahima, T., Röder, M. S., Li, Y., Nevo, E., et al. (2003). Domestication quantitative trait loci in triticum dicoccoides, the progenitor of wheat. *Proc. Natl. Acad. Sci. U. S. A.* 100(5), 2489–2494. doi: 10.1073/pnas.252763199
- Pritchard, J. K., Stephens, P., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. 155(2), 945–959. doi: 10.1093/genetics/155.2.945
- Quamruzzaman, M., Manik, S. M. N., Shabala, S., Cao, F., and Zhou, M. (2021). Genome-wide association study reveals a genomic region on 5AL for salinity tolerance in wheat. *Theor. Appl. Genet.* 135(2), 709–721. doi: 10.1007/s00122-021-03996-8
- Rahimi, Y., Bihamba, M. R., Taleei, A., Alipour, H., and Ingvarsson, P. K. (2019). Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. *BMC Plant Biol.* 19(1), 149. doi: 10.1186/s12870-019-1754-6
- Rajaram, S., Van Ginkel, M., and Fischer, R. A. (1994). CIMMYT's wheat breeding mega-environments. *Proceedings of the 8th International Wheat Genetics Symposium*, Beijing, China.
- R Development Core team (2021). R: a language and environment for statistical computing. *Version 4.0.5. Vienna, Austria: R Foundation for Statistical Computing*; <https://www.R-project.org/>. R. A. Lang. Environ. Stat. Comput. R Found. Stat. Comput. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- Roncallo, P. F., Akkiraju, P. C., Cervigni, G. L., and Echenique, V. C. (2017). QTL mapping and analysis of epistatic interactions for grain yield and yield-related traits in triticum turgidum l. var. durum. *Euphytica*. 213, 277. doi: 10.1007/s10681-017-2058-2
- Roncallo, P. F., Beaufort, V., Larsen, A. O., Dreisigacker, S., and Echenique, V. (2018). Genetic diversity and linkage disequilibrium using SNP (KASP) and AFLP markers in a worldwide durum wheat (*Triticum turgidum* l. var durum) collection. *PLoS One*. 14, 1–33. doi: 10.1371/journal.pone.0218562
- Rufo, R., Alvaro, F., Royo, C., and Soriano, J. M. (2019). From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. *PLoS One*. 14(7), e0219867. doi: 10.1371/journal.pone.0219867
- Sabadin, P. K., Malosetti, M., Boer, M. P., Tardin, F. D., Santos, F. G., Guimarães, C. T., et al. (2012). Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences. *Theor. Appl. Genet.* 124(8), 1389–402. doi: 10.1007/s00122-012-1795-9

- Sall, A. T., Chiari, T., Legesse, W., Seid-Ahmed, K., Ortiz, R., Van Ginkel, M., et al. (2019). Durum wheat (*Triticum durum* desf.): Origin, cultivation and potential expansion in sub-saharan Africa. *Agronomy*. 9(5), 263. doi: 10.3390/agronomy9050263
- Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petrolini, C., et al. (2020). Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat. Commun.* 11(1), 4572. doi: 10.1038/s41467-020-18404-w
- Shi, J., Habben, J. E., Archibald, R. L., Drummond, B. J., Chamberlin, M. A., Williams, R. W., et al. (2015). Overexpression of ARGOS genes modifies plant sensitivity to ethylene, leading to improved drought tolerance in both arabidopsis and maize. *Plant Physiol.* 169(1), 266–82. doi: 10.1104/pp.15.00780
- Siol, M., Jacquin, F., Chabert-Martinello, M., Smýkal, P., Le Paslier, M. C., Aubert, G., et al. (2017). Patterns of genetic structure and linkage disequilibrium in a large collection of pea germplasm. *G3 Genes Genomes Genet.* 7(8), 2461–2471. doi: 10.1534/g3.117.043471
- Soriano, J. M., Villegas, D., Aranzana, M. J., García Del Moral, L. F., and Royo, C. (2016). Genetic structure of modern durum wheat cultivars and mediterranean landraces matches with their agronomic performance. *PLoS One*. 11(8), e0160983. doi: 10.1371/journal.pone.0160983
- Soriano, J. M., Villegas, D., Sorrells, M. E., and Royo, C. (2018). Durum wheat landraces from east and west regions of the mediterranean basin are genetically distinct for yield components and phenology. *Front. Plant Sci.* 8(80), 40. doi: 10.3389/fpls.2018.00080
- Sukumaran, S., Reynolds, M. P., and Sansaloni, C. (2018). Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. *Front. Plant Sci.* 9, 81. doi: 10.3389/fpls.2018.00081
- Sun, F., Xu, M., Park, C., Dwiyantri, M. S., Nagano, A. J., Zhu, J., et al. (2019). Characterization and quantitative trait locus mapping of late-flowering from a Thai soybean cultivar introduced into a photoperiod-insensitive genetic background. *PLoS One*. 14(12), e0226116. doi: 10.1371/journal.pone.0226116
- Suprayogi, Y., Pozniak, C. J., Clarke, F. R., Clarke, J. M., Knox, R. E., and Singh, A. K. (2009). Identification and validation of quantitative trait loci for grain protein concentration in adapted Canadian durum wheat populations. *Theor. Appl. Genet.* 119(3), 437–48. doi: 10.1007/s00122-009-1050-1
- Talini, R. F., Brandolini, A., Miculan, M., Brunazzi, A., Vaccino, P., Mario Enrico, P., et al. (2020). Genome-wide association study of agronomic and quality traits in a world collection of the wild wheat relative *triticum urartu*. *Plant J.* 102(3), 555–568. doi: 10.1111/tpj.14650
- Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). GAPIT version 2: An enhanced integrated tool for genomic association and prediction. *Plant Genome*. 9(2). doi: 10.3835/plantgenome2015.11.0120
- Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Front. Physiol.* 3, 347. doi: 10.3389/fphys.2012.00347
- Turki, N., Shehzad, T., Harrabi, M., and Okuno, K. (2015). Detection of QTLs associated with salinity tolerance in durum wheat based on association analysis. *Euphytica*. 201, 29–41. doi: 10.1007/s10681-014-1164-7
- Tzarfati, R., Barak, V., Krugman, T., Fahima, T., Abbo, S., Saranga, Y., et al. (2014). Novel quantitative trait loci underlying major domestication traits in tetraploid wheat. *Mol. Breed.* 34, 1613–1628. doi: 10.1007/s11032-014-0182-4
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11), 4414–4423. doi: 10.3168/jds.2007-0980
- Varshney, V., and Majee, M. (2022). Emerging roles of the ubiquitin – proteasome pathway in enhancing crop yield by optimizing seed agronomic traits. *Plant Cell Rep.* 41(9), 1805–1826. doi: 10.1007/s00299-022-02884-9
- Velu, G., Tutus, Y., Gomez-Becerra, H. F., Hao, Y., Demir, L., Kara, R., et al. (2017). QTL mapping for grain zinc and iron concentrations and zinc efficiency in a tetraploid and hexaploid wheat mapping populations. *Plant Soil*. 8, 680391. doi: 10.1007/s11104-016-3025-8
- Walter, S., Kahla, A., Arunachalam, C., Perochon, A., Khan, M. R., Scofield, S. R., et al. (2015). A wheat ABC transporter contributes to both grain formation and mycotoxin tolerance. *J. Exp. Bot.* 66(9), 2583–2593. doi: 10.1093/jxb/erv048
- Wang, S., Xu, S., Chao, S., Sun, Q., Liu, S., and Xia, G. (2019). A genome-wide association study of highly heritable agronomic traits in durum wheat. *Front. Plant Sci.* 10, 919. doi: 10.3389/fpls.2019.00919
- Wang, J., and Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinf.* 19(4), 629–640. doi: 10.1016/j.gpb.2021.08.005
- Wanke, D., and Üner Kolukisaoglu, H. (2010). An update on the ABC transporter family in plants: Many genes, many proteins, but how many functions? *Plant Biol.* 12(1), 15–25. doi: 10.1111/j.1438-8677.2010.00380.x
- Weir, B. S. (1997). Genetic data analysis II. *Biometrics*. doi: 10.2307/2533134
- Wu, H., Ni, Z., Yao, Y., Guo, G., and Sun, Q. (2008). Cloning and expression profiles of 15 genes encoding WRKY transcription factor in wheat (*Triticum aestivum* L.). *Prog. Nat. Sci.* 18(6), 697–705. doi: 10.1016/j.pnsc.2007.12.006
- Xynias, I. N., Mylonas, I., Korpetis, E. G., Ninou, E., Tsballa, A., Avdikos, I. D., et al. (2020). Durum wheat breeding in the Mediterranean region: Current status and future prospects. *Agronomy*. 10(3), 432. doi: 10.3390/agronomy10030432
- Yu, X., Jiang, Y., Yao, H., Ran, L., Zang, Y., and Xiong, F. (2021). Cytological and molecular characteristics of delayed spike development in wheat under low temperature in early spring. *Crop J.* 10(3), 840–852. doi: 10.1016/j.cj.2021.08.008
- Zhang, L., Luo, J. T., Hao, M., Zhang, L. Q., Yuan, Z. W., Yan, Z. H., et al. (2012). Genetic map of triticum turgidum based on a hexaploid wheat population without genetic recombination for d genome. *BMC Genet.* 13, 69. doi: 10.1186/1471-2156-13-69
- Zhao, M. M., Zhang, X. W., Liu, Y. W., Li, K., Tan, Q., Zhou, S., et al. (2020). A WRKY transcription factor, TaWRKY42-b, facilitates initiation of leaf senescence by promoting jasmonic acid biosynthesis. *BMC Plant Biol.* 20(1), 444. doi: 10.1186/s12870-020-02650-7