Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E ALIMENTARI

Ciclo 35

Settore Concorsuale: 07/E1 - CHIMICA AGRARIA, GENETICA AGRARIA E PEDOLOGIA

Settore Scientifico Disciplinare: AGR/07 - GENETICA AGRARIA

GWAS ANALYSIS FOR THE IDENTIFICATION OF MOLECULAR BASIS RESPONSIBLE FOR YIELD RELATED TRAITS AND DISEASE RESISTANCE IN DURUM WHEAT

Presentata da: Matteo Bozzoli

Coordinatore Dottorato

Massimiliano Petracci

Supervisore

Marco Maccaferri

Co-supervisore

Roberto Tuberosa

Cristian Forestan

Esame finale anno 2023

Contents

1. Summary of the thesis

The PhD thesis was elaborated in the framework of the Innovar H2020 project "Next generation variety testing for improved cropping on European farmland" (2019-2024). The Innovar project is a European project which involves several research institutes and breeding companies across different European nations. The principal aim of the project is to innovate the procedure of wheat varietal registration using genomics, phenomics and machine learning technologies. Currently, new wheat varieties, to be included in European Union common catalogue and be awarded with the Plant Variety Protection (PVP) system, must follow the criteria included in the distinctness, uniformity and stability (DUS) protocols for at least two years of field trials based on morphological traits to describe plant development and agronomic features. Furthermore, additional field trials are required to establish the value of cultivation and use (VCU) using EU protocols, which is more focused on yield related traits. The VCU tests are only required for the inscription of the new variety to the national register of varieties, which are then automatically included in the European common catalogue by communication of each state member. Once varieties are included in the EU common catalogue, they can be commercialized across European nations.

Currently, these EU protocols rely only on phenotypic characteristic and morphological traits, strongly influenced by the environment and by visual assessment. One of the main objectives of Innovar project is to use genomics data on durum and bread wheat panels to associate molecular markers and candidate genes on the DUS and VCU traits, augmenting the information in the EU protocol. By associating the phenotypic traits to molecular markers or genomic regions (QTLs), the work by breeders to register newly bred varieties will be facilitated, as markers strongly associated to the phenotypic traits will be available for marker assisted selection analysis (MAS) and molecular breeding. The herein reported PhD thesis is mainly developed on the Innovar project, structured in three main chapters dealing with: 1) GWAS analysis on durum VCU and DUS traits related to yield and morphological agronomic traits, 2) resistance of durum wheat modern varieties to soil borne cereal mosaic virus, 3) study of the spike fertility trait in segregant durum population, in connection with the DUS protocol.

The first chapter summarizes the results obtained by Innovar project regarding durum wheat trials, using genomics and phenotyping data to carry out genome wide association studies (GWAS) to obtain associated QTLs on different agronomic traits. The Innovar project include several field trials across Europe for bread and durum wheat, which were phenotyped following the VCU and DUS protocols. Herein it is reported the work performed on VCU and DUS durum panels, by analysing data across different locations and performing GWAS analysis. The Innovar project divides the VCU durum trials between different agro-climatic zones (ACZ): Mediterranean, Maritime South and Pannonian. The Mediterranean ACZ is composed by two environments for durum trials (UPM and CSIC partners), Maritime South ACZ is composed by other two environments (HORTA and UNITUS), Pannonian is composed by one environment (UNIDEB). The VCU core trials (166 varieties) were divided in these ACZ in each field trial sowing 5 varieties common to all ACZ, 5 varieties specific for each ACZ and 35 varieties specific for each field trial. Furthermore, two fungicide managements (full and minimal treatments) were applied with wide spectrum chemicals, using three replicates for each treatment. In addition, the durum DUS trial (253 varieties, composed by VCU durum panel and Unibo background material to increase genetic diversity) was performed in collaboration with CREA-DC (Italy), sowing genotypes in two replicates following the alpha-design randomization. To

summarize, during the first year growing season, VCU phenotypic data were obtained across the different field trials as well as the DUS morphological traits in only one environment (CREA-DC). The VCU and DUS phenotypic data were statistically analysed with Rstudio software to obtain the best linear unbiased estimations (BLUEs). In addition, as regards the VCU, the interaction between genotypes and environments was estimated and included in the mixed model to extract BLUEs, to correct phenotypic values for the environment and treatment influence. The whole DUS durum panel (which included also VCU varieties) was genotyped using Infinum Illumina 90K SNPChip (Wang et al., 2014), augmented with genotypic data from global durum resources in Unibo, such as the global durum panel (GDP) (Mazzucotelli et al., 2020), the durum panel 1 (DP1) and durum panel 2 (DP2) (Maccaferri et al., 2015). The augmented panel was created on order to increase genetic diversity by including modern durum varieties coming from different subpopulation origins and breeding program. The genotypic data were filtered for quality traits, retaining only SNPs correctly mapped based on the consensus map with position on *Triticum turgidum* Svevo RefSeqv1.0 pseudomolecule (Maccaferri et al., 2019a, 2015). The population structure and genetic similarities were calculated for the whole durum genotypes, detecting specific groups of subpopulations for the different breeding programs but also a high percentage of admixed lines, meaning that a high number of genotypes share the same genetic background, breeding origins or pedigree founders.
Once the genotypic and phenotypic data were obtained, genome wide association analysis (GWAS) was performed on VCU and DUS durum datasets, considering kinship matrix (genetic similarities) to correct for population structure. As regards VCU, yield related traits were mainly considered such as heading date, yield corrected for 15% of moisture content, plant height and thousand grain weight (TGW). In addition to the VCU trials, the DUS trials were analysed with GWAS for all the different characters, showing interesting peaks for plant growth habit, heading date, ear and awn glaucosity, ear and awn colour. In particular, as regards heading date, the *Ppd-2B* locus on the chromosome 2B was detected (Maccaferri et al., 2008). In addition, candidate gene networks were enquired using Knetminer database to understand gene functions and metabolomic areas. In general, for yield related traits in VCU, principal metabolic functions are ascribed to flowering time, photoperiod regulation, inflorescence architecture, but also yield traits in relation to grain number and size. To conclude, the Innovar project represents a main part of the PhD thesis where interesting results have been reached regarding durum wheat genetics. In fact, the exploitation of the genetic similarities between the durum panel accessions revealed the relationships between each genotype that will be used to determine genetic similarity thresholds (GSV). In fact, InnoVar will calculate the genetic similarity value threshold to use as reference to discriminate varieties that can be selected to be grown alongside candidate varieties to determine their distinctness (D). The determination of the GSVs will help in the identification of the best performing varieties in different environments and to propose genetically different varieties in the framework of the Innovar project. Furthermore, different QTLs were detected for the main DUS and VCU traits detecting molecular markers strongly connected with the main phenotypic traits that could be used to augment the variety registration protocols and the marker assisted selection.

The DUS and VCU durum field trials are mainly focused on yield related traits and morphological agronomical trait, while disease is only partially considered. In Unibo, durum wheat resistance to soil borne cereal mosaic virus (SBCMV) was considered in order to augment the information on durum wheat varieties included in the Innovar project. The background work performed in Unibo relies on the mapping of the s*bm2* QTL on short arm of chromosome 2B in segregant mapping

population (Maccaferri et al., 2011a). Bruschi et al. (*unpublished*), confirmed the presence of the QTL on the Unibo durum panel 1 (DP1) performing GWAS analysis with genotypic data from Illumina 90K SNP Chip (Wang et al., 2014a) and phenotypic data from multi-year experiment collected in Unibo SBCMV infected field. Furthermore, the *sbm2* interval was fine mapped using RILs segregant population Svevo (R) x Ciccio (S) and Meridiano (R) x Claudio (S) by converting 11 SNPs to KASP markers, reaching and interval of 3 Mb (from 13 Mb to 16 Mb on chromosome 2BS). In this PhD theis, *sbm2* interval was further fine mapped using the backcross population Meridiano x MeridianoClaudio BCF_4 (MxMC110). Critical recombinant lines were detected in multi-years field trials in Unibo SBCMV field thanks to further 6 KASP markers that were developed from Affimetrix 420K SNP Chip array. As a consequence, the *sbm2* interval was fine mapped to 1.1 Mb between 14.7 Mb and 15.8 Mb on chromosome 2B. The fine mapped interval included 14 high confidence genes (HC) potentially involved in resistance reaction, such as protein kinases that could be involved in signal transduction, *NBS-LRRs* involved in specific resistance reaction and different secondary metabolic pathways. In order to confirm the genetic interval, a comparison was performed with the second version of the Svevo genome, Svevo Platinum (Unibo background data, *unpublished*). The comparison of the *sbm2* interval between the two genomes showed an inversion in the QTL interval annotated on Svevo RefSeq v1.0 genome, between 14.3 Mb and 14.8 Mb on chromosome 2B. The inversion is in a critical recombination point for the fine mapped interval of 1.1 Mb and was considered when evaluating candidate genes on the interval. In order to better explore the *sbm2* interval, an RNAseq experiment was performed on elite durum cultivars resistant or susceptible to the SBCMV based on haplotypic data and background information. During 2021 growing season, RNAseq was performed collecting roots in a time course experiment (three dates), sequencing resistance bulks (Meridiano, Levante and Neodur), susceptible bulk (Simeto, Claudio and Altar-84), single resistant and susceptible varieties (Svevo and Ciccio). The differential expressed genes were obtained from RNAseq reads, clusters of genes overexpressed in resistance cultivars were identified with particular interest on three genes strongly involved in candidate gene functions: *TRITD2Bv1G007390* (protein kinases), *TRITD2Bv1G007240* (NBS-LRR) and *TRITD2Bv1G007400* (protein kinases). These cluster of differentially expressed genes were confirmed repeating the RNAseq analysis on Svevo Platinum transcriptome, after shifting the annotation from Svevo RefSeq v1.0 published assembled genome. The analysis of RNAseq data on Svevo Platinum transcriptome showed two genes strongly overexpressed in resistant cultivar at different time points, in comparison to the susceptible ones: *TRITD2Bv1G007260* and *TRITD2Bv1G007390,* two protein kinases.

The next question to address was to clarify if the SBCMV was recognized for the resistance reaction or if it was the fungal vector as well. In order to understand this issue, the unmapped reads from the previous RNAseq experiment were mapped on the SBCMV RNA1-2 and on the *Polymixa beate* genome. The results showed that there is a higher number of reads from susceptible samples that map on SBCMV RNA1-2 and that there is not significant difference between resistant and susceptible cultivars mapping the RNAseq reads on the *Polymixa beate* genome. This proves that the SBCMV seems to be the only responsible for the plant resistant reaction, recognized in the plant root tissue, in comparison to vector fungus *Polymixa graminis* which is completely absent from host plant radical tissue.

Furthermore, a haplotype analysis was performed, using 4 KASP markers developed from Illumina 90K SNPs. The durum panel 1 (DP1) was augmented with the Innovar durum panel reaching a final number of more than 500 varieties. The genotyping analysis showed resistance, susceptible and

recombinants haplotypes between these KASP markers tagging the haplotype. Further analysis on these genotypes were performed considering the different breeding origins and different years of varietal inscription, showing that resistance sources to SBCMV gradually decreased in recently registered varieties. The herein reported results on the resistance to SBCMV highlight important candidate genes that could be responsible for the resistant haplotypes of durum varieties. This strengthens the importance of detecting the candidate genes and pyramid the QTLs both in bread and durum wheat varieties under registration procedure, trying to counteract the SBCMV spread that depletes yield production about 70% in Europe.

Connected with the Innovar project, in addition to the DUS morphological traits considered, spike fertility trait was studied in Unibo starting from the multi-cross population Neodur-Claudio-Colosseo-Rascon/2*Tarro (NCCR) (Milner et al., 2016). The work from Milner and colleagues was the background material which was used to further characterize the spike fertility trait. Basically, the NCCR population was phenotyped for the number of florets per spikelet and performed GWAS using Illumina Infinium Array 90K SNP Chip, finding a strong QTL on chromosome 2A. This QTL was responsible for grain number increase and was named *GNI-2A,* mapped on chromosome 2A with an R^2 value of 0.42 in the NCCR population. The *GNI-2A* haplotype, paralogue of *GNI-1* (orthologue of *vrs1* in barley) (Sakuma et al., 2019, 2017), was responsible for an higher number of florets per central spikelet hampering florets abortion. The *GNI-2A* haplotype was found only in Rascon/2*Tarro (CYMMIT germplasm derived from Altar-84) and not in others parental lines, which showed the spike fertility phenotypic trait. Starting from the most associate SNP marker, the confidence interval was identified considering +/- log=2, and six SNPs were converted to KASP markers (from KUBO 44 to KUBO 50). The KASP markers were used to genotype the durum panel 1 (DP1) accessions (167 genotypes), the *GNI-2A* was identified as a rare allele (frequency percentage of 2.4%) and in CYMMIT related varieties.
The spike fertility trait was assessed also in the segregant biparental population Relief x Iride F_6 (1500 genotypes), which was phenotype for number of florets per central spikelets in two consecutive seasons. The phenotype distribution of the trait is represented by a bimodal curve, which segregates for the number of florets per central spikelet and with high heritability values (0.89). Based on the confidence interval KASP markers were used to genotype the whole Relief x Iride population, confirming the confidence interval of *GNI-2A* between markers *KUBO 44* and *KUBO 49* of about 5 Mb. Furthermore, the genotypes carrying the wild-type and *GNI-2A* haplotypes detected in Relief x Iride population, based on KASP analysis, showed strong phenotypic differences between the two groups, and 10 recombinants in the confidence interval were detected. In order to further fine map the confidence interval, further 6 KASP markers were converted from Affimetrix 420K SNP Chip array and used to genotype the 10 informative recombinants. The spike fertility characteristics were calculated for the three central spikelets in 6 spikes per genotype, collected from the 2020 field nursery. Merging phenotypic and genotypic data for the 10 recombinants, the confidence interval was narrowed down to 3.9 Mb.
The fine mapped confidence interval contains around 10 high confidence genes, but the strongest candidate is represented by *hox2*, an homeobox leucin zipper transcription factor paralogue of GNI-1 (Sakuma et al., 2019). By background resequencing analysis performed on different durum cultivars in collaboration with IPK (Gatersleben, Germany), some of which have the *GNI-2A* haplotype, the gene seem to have a 4 kbp deletion in increased spike fertility varieties. This deletion in *hox2* gene seems to be responsible for the increase spike fertility trait as it is not present in wt

accessions, such as Svevo or Relief. To conclude, *hox2* paralogue *GNI-1* has a single amino acidic mutation that prevents from the inhibition of apical florets abortion (Sakuma et al., 2019), this function can be connected with the *hox2 GNI-2A* mutation on Altar-84 related genotypes (Rascon/2*Tarro and related CYMMIT germplasm), as the genes carrying the 4 kbp deletion lead to an increase in spike fertility preventing florets abortion.

2 Introduction

2.1 Origin and development of durum wheat

Cereals represent an important crop for worldwide food supply and population sustainability, being of major interest for both population market and animal feed (http://www.fao.org/DOCREP/006/Y4683E/y4683e06.htm#TopOfPage). Wheat is considered as one of the major crops among cereals, where more than 600 million of tonnes are harvested every year worldwide (Shewry, 2009). The genus wheat (*Triticum* spp.) belongs to the *Poaceae* family, (*Triticae* tribe and *Triticineae* subtribe), which includes annual grass. The wheat genus is divided in two mainly relevant crops, which are widely distributed in agricultural market for feeding purposes, namely: the hexaploid bread wheat (*Triticum aestivum*) and the tetraploid durum wheat (*Triticum turgidum* subsp. *durum*). In comparison to other cereal crops, wheat has a very wide range of cultivation and diversity, from norther Scandinavian region to tropics and southern America ("Evolution of crop plants /," 1995). In particular, durum wheat is more adapted to Mediterranean or rainfed and semi-arid climate zones, cultivated in temperate climates and principally used by pasta industry (Shewry, 2009). About 95% of cultivated wheat in the world is represented by common wheat (bread wheat), principally used to make bread, cookies and pastries, whilst the remaining 5% is represented by durum wheat principally used to make pasta and other related products (Dubcovsky and Dvorak, 2007). Beside bread and durum wheat, other species, such as einkorn, emmer and spelt, are also cultivated in Mediterranean and mid-oriental regions but their production is limited to specific areas / markets (Abdel-Aal et al., 1998; Szabo and Hammer, 1996; Dubcovsky and Dvorak, 2007).

Durum and bread wheat are allopolyploids, the first is tetraploid (AA and BB genomes) and the second one is hexaploidy (AA, BB and DD). The polyploidy occurred because of natural hybridization events in Fertile Crescent and Transcaucasian territories. Domestication and human selection produced crops more adapted to farming procedures (Dubcovsky and Dvorak, 2007).

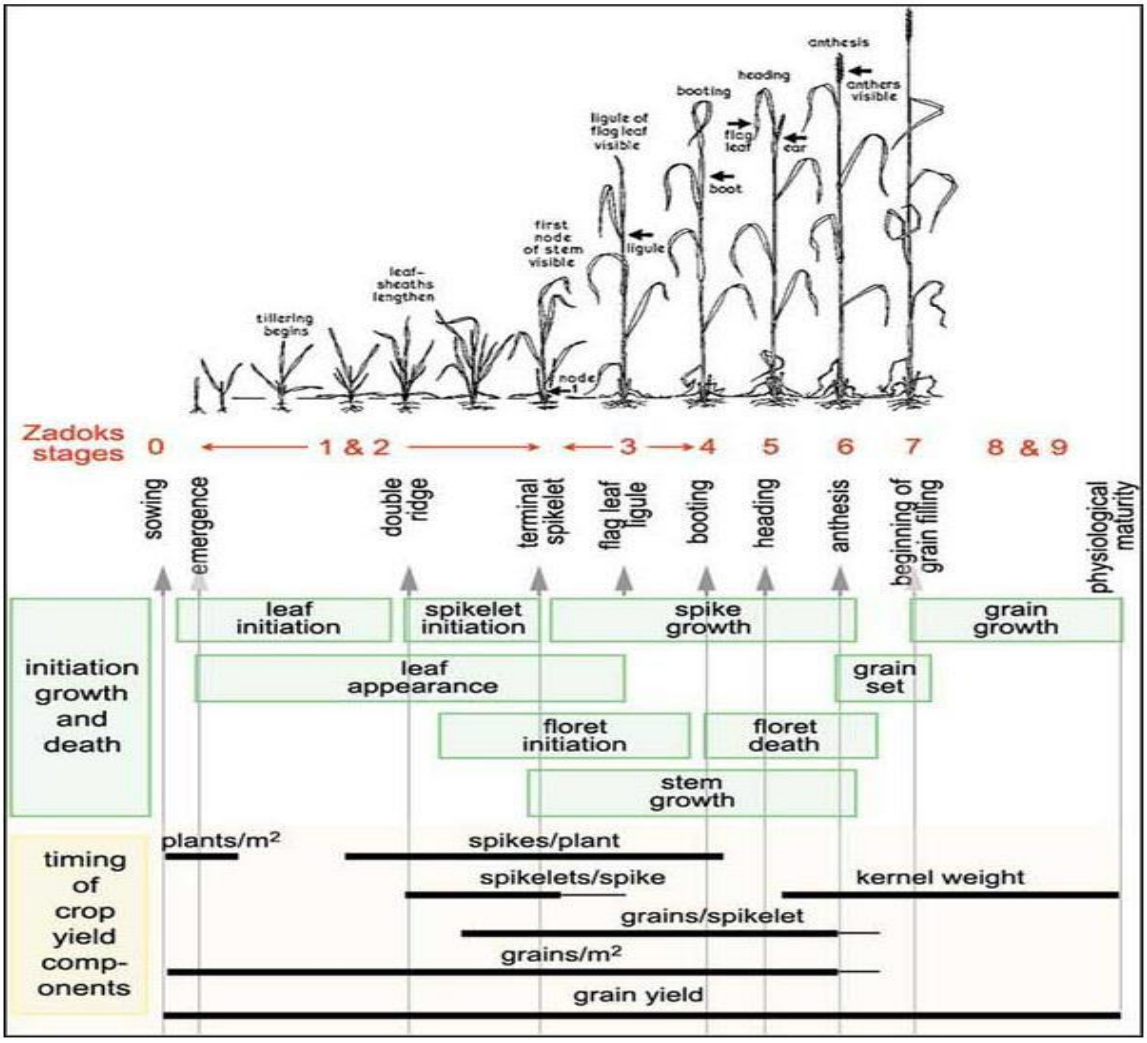
The evolution and domestication process of wheat occurred 10,000 years ago as part of the Neolitic revolution, where agriculture gained more importance for food supply in the Fertile crescent region (Maccaferri et al., 2019b; Shewry, 2009). The first hybridization event occurred between *Triticum urartu* (AA genome) and a progenitor of the BB genome (an Ancient Aegilops from the *Sitopsis* branch), originating the tetraploid wild emmer wheat (AABB) (Heun et al., 1997) about 0.5 million years ago, known also as wild emmer wheat (WEW) (Maccaferri et al., 2019b). As regards to the B genome, its origin is debated and more complex to be confirmed ("Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin | Oxford Academic," n.d.). It has been hypothesised that the BB genome comes from a similar species to the today known *Aegilops speltoides*, originating from the SS genome ((Sarkar and Stebbins, 1956). The process of domestication, from wild emmer wheat, produced the domesticated emmer wheat (DEW) which then lead to durum wheat landraces and then to durum wheat modern cultivars (DWC) (Maccaferri et al., 2019b). The generation of domesticated modern wheat has the first evidences back to 7,000 years ago, even if the widespread of the crop occurred just 2,000 years ago (Faris et al., 2014). Finally, the hybridization that generated modern bread wheat (AABBDD, hexaploid) occurred between tetraploid wheat (AABB) and *Aegilops tauschii* (DD genome) in the Caspian sea region (the TransCaucasian corridor, Wang et al., 2013). The modern durum wheat AABB genome was shown to be homologue and interfertile with the wild

and domesticated emmer wheat (Akhunov et al., 2010). On the other hand, the bread wheat hexaploidy genome (AABBDD) show similarities with *Triticum aestivum* ssp. spelta, i.e. spelt (Slageren, 1994). Based on that, wild species represent an important source of genetic diversity for interesting agronomic characteristics (such as disease resistance or spike architecture traits). This stress the importance of comparative genomics analysis between modern wheat species and wild ancestors in order to detect beneficial genomic regions or alleles to be introgressed in modern cultivars by breeding programs (Brozynska et al., 2016; Gaut, 2015).

The loss of genetic diversity and presence of favourable alleles in modern wheat germplasm, is due to the domestication process which separated modern bred cultivars originated from few founders to the wild progenitors (Shewry, 2009). The selective pressure was performed year by year on some key agronomic characters on wild progenitors, such shattering of spike at maturity which caused severe yield losses in wheat ancestors (Dubcovsky and Dvorak, 2007). The loss of this trait, important for natural dissemination but negative for human exploitation is regulated by mutations on the *Br* locus (Nalam et al., 2006). Another important trait subjected to selection in modern wheat was the free threshing glumes, selecting cultivars where glumes did not adhere to the grain facilitating the threshing process. This glume form is regulated by the *Q* locus. Human selected negatively the wild allele correspondent to tenacious glumes (*Tg*) locus and selected mutants with free-threshing phenotypes (Jantasuriyarat et al., 2004; Simons et al., 2006; Dubcovsky and Dvorak, 2007). Additional important traits selected for domestication includes seed size, decreased tiller number, erect growth habit and reduced seed dormancy (Dubcovsky and Dvorak, 2007).

The domestication process, together with the hallopolyploidization, allows especially bread wheat to be more adapted to different environments and wide range of conditions (Dubcovsky and Dvorak, 2007). This is especially true for bread wheat AABBDD genome, in comparison to durum wheat (AABB).

Durum wheat breeding of modern cultivars is focused on selecting interesting agronomic traits to define better performing varieties. The modelling of wheat development relies on the different phenological stages of wheat, as a results of the interaction between genetic and environmental factors (Landsberg, 1977). Wheat is divided in time scale growth stages, namely: germination, seedling growth, tillering, stem elongation, booting, ear emergence, flowering, milk development, dough development and ripening. These stages can be grouped in three main phases (Slafer and Rawson, 1994): 1) vegetative stage, from seed imbibition to floral initiation 2) early reproductive and 3) late reproductive/maturity stage (from terminal spikelet initiation to end of grain filling). A commonly used scale to discriminate wheat growth stages was reported by Zadoks (Zadoks et al., 1974), which ranges from 0 to 99 (Figure 1). In particular, these two digits scheme reports the first digit as the principal wheat growth stages, whilst the second digit reports the stage of development. It is important to use such developmental references and growing scales in any phenotyping activity.



.

Figure 1: growth scale of wheat based on Zadoks stages. The scale is divided between growth stages and the corresponding times of plant components

2.2 Durum wheat breeding

The history of durum wheat breeding has always had the main purposes to increase yield production and resistance to biotic and abiotic stresses. The precursor of durum wheat breeding was Strampelli who performed selection on southern Italian landraces releasing the widely known cultivar "Senatore Cappelli" in 1920, representing an important cultivated variety in Italy but also in the Mediterranean region ("Options Méditerranéennes en ligne - Collection numérique - Evolution of durum wheat breeding in Italy," n.d.). Breeding programs and varietal selection lead to a significant increase in yield in the second part of 20th century as a consequence of the "Green revolution", where the breeding for high production led to the development of semi dwarf cultivars more responsive to nitrogen fertilizers (for higher N uptake) and resistant to lodging (Hedden, 2003; Isidro et al., 2011). It is worth mentioning that the CIMMYT and ICARDA programs played an important role in durum variety registration with an high percentage of varieties that derived from CIMMYT parental lines, selecting for cultivars with increasing yield production and favourable agronomic traits (Bassi et al., 2019; Pfeiffer et al., 2000). In particular, CIMMYT breeding programs selected for

varieties with increased number of grains per spike which led also to higher number of grains m^2 (Pfeiffer et al., 2000; Waddington et al., 1987). This trait was transferred also to Italian and Spanish cultivars in late 90's corresponding to a genetic gain for modern bred cultivars without changing the total number of spikelets per spike (Royo et al., n.d.; Vita et al., 2007). In addition to yield, another related trait regards the grain quality (De Vita et al., 2007;Motzo et al., 2004). Genetic improvement related to increasing yield component is negatively correlated to protein content due to the dilution of proteins as a consequence of increased grain size and starch accumulation. However, this did not affect pasta quality, the main application of durum wheat, as modern cultivars were selected to have an increased gluten index (De Santis et al., 2017; Guarda et al., 2004; Motzo et al., 2004) and in some cases also protein content.

Breeding programs use different methods for variety selection, as for example the pedigree method, backcross, single seed descent or pure line selection (Xynias et al., 2020). All these methods were used in durum wheat breeding to focus on grain yield and quality, adaptation to environment of newly bred cultivars, efficient use of resources and resistance to biotic and abiotic stresses (Xynias et al., 2020).

The use of molecular markers in breeding programs represent a major advantage to speed up the selection processes in breeding programs. Molecular markers are genomic sequences (or even single alleles) that are conserved in genomes and are amenable to develop a laboratory assay to genotype individuals and breeding lines/materials.

Molecular markers can be used to estimate genetic variability in germplasm collections of genotypes and can be associated (by linkage) to important agronomic traits subject of breeding selection (Abu-Zaitoun et al., 2018; Achtar et al., 2010). The use of molecular markers has different advantages in plant breeding: selection can occur in early developmental stages; the backcross can be accelerated and the introgressed alleles from wild species in modern cultivars can be monitored. This can be performed using molecular markers that are strongly associated with the agronomic trait of interest where an allelic variation of the molecular marker is responsible for the phenotypic variance, identifying quantitative traits loci (QTLs) (Kucek et al., 2015). QTLs are identified as genomic regions associated with the genetic variation of complex traits (Geldermann, 1975). The study of a QTL is associated with the phenotypic expression of determined agronomic traits, where molecular markers are statistically associated with phenotypic variance (Lynch et al., 2017). These markers are used for marker assisted selection (MAS) which can be performed by phenotyping a genetic mapping population followed by QTL analysis (Kucek et al., 2015).

Different QTL studies were conducted in durum wheat breeding programs, to dissect the genetic inheritance of important agronomic traits using molecular markers and different genetic maps, such as grain yield traits, grain protein content and pasta quality related characteristics (Maccaferri et al., 2008; Peleg et al., 2009). Other traits dissected by QTL mapping regards biotic stresses to select varieties based on fungi resistance (Prat et al., 2017) and pests (Varella et al., 2019), or also abiotic stresses such as heat and drought resistance (El Hassouni et al., 2019).

Another objective of breeding was to delay or shorten the time to anthesis increasing the period of photosynthetic activity, mainly achieved with the deployment of diverse alleles at major genes responsible for the photoperiod sensitivity (*Ppd*) (Foulkes et al., 2004; Turner et al., 2005) and vernalization genes (*Vrn*) (Yan et al., 2006, 2003).

The QTL analysis reached the maximum application with the development of next generation sequencing technologies (NGS), with the exploitation of single nucleotide polymorphisms (SNPs). These molecular markers represent high informative single nucleotide alleles mutations which are consistent and widespread in the genome, different genotyping arrays were developed to perform high throughput phenotyping of wheat panels or mapping populations, such as Illumina 90K SNP Chip (Wang et al., 2014b). The 90K SNP Chip was used also to produce different genetic maps, as for

example the cross derived from wild emmer and modern durum wheat (Raz Avni et al., 2014) or others published by Maccaferri et al (2015).
Another application of high throughput arrays is to perform genomic selection to identify the genetic value of lines under selection by predicting the genotype agronomic characteristics starting from genotypic data (Newell and Jannink, 2014). By using high throughput genotypic data, genomic selection develops models to capture all QTLs also with minor effect on the population, without removing them in the selection steps performed in the different breeding methods.

2.3 QTL dissection and GWAS analysis

As already mentioned, next generation sequencing technology and the development of SNP Chip arrays played an important role in mapping approaches to study QTL segregation in experimental populations (Alqudah et al., 2020). Examples of genotyping arrays can be the 9K iSelect Illumina array (Comadran et al., 2012), the 50K iSelect Infinum Illumina array (Bayer et al., 2017) or the 90K iSelect Infinum Illumina array (Wang et al., 2014b). The association between phenotype and genomic regions is studied with two different methods: QTL mapping using biparental populations, or alternatively Genome-Wide Association mapping (GWAS) using genotype panels of unrelated individuals.
QTL mapping is usually used to dissect genomic regions that control target traits, different types of mapping populations could be used for this purpose, such as double haploids or recombinant inbred lines (RILs) (Alqudah et al., 2020). However, one of the major limitations of using biparental populations to dissect QTLs is that only segregating alleles between the two parental lines can be studied (Mitchell-Olds, 2010). In this case, a possible solution could be to perform multi-parental crosses for increasing recombination and allele segregations and augment the number of informative markers (Liller et al., 2017).
As regards GWAS, the analysis relies on association population composed of accessions which are not related to each other to a large extent and where different recombination accumulated over time and with alleles maintained over generations as a consequence of linkage disequilibrium (LD), which represents the association of molecular markers that are inherited together over generations (Alqudah et al., 2020). The GWAS analysis has an high probability to detect QTL regions responsible for phenotypic variations if high density molecular SNP arrays are used, such as the 90K SNP consensus map which is represented by correctly mapped SNP using durum genetic maps (Maccaferri et al., 2015). The GWAS analysis has some key aspect to be considered in order to have consistent and reliable data. First of all, the phenotypic data need to be correctly analyzed, removing outliers from the phenotypic data and calculating heritability value of the traits, which ranges from 0 (low) to 1 (high) and corresponds to the value of genetic component responsible for phenotypic variation. To evaluate the environment effect that may deplete heritability values in multi-years or multi-enviromental trials, the standardization of phenotypic value can occur calculating the best linear unbiased prediction (BLUPs) or best linear unbiased estimator (BLUEs) (Alqudah et al., 2020).
Another key aspect to be considered is the number of individuals. Using an high number of individuals for GWAS analysis increase the opportunity to have significant associations between the molecular marker and the phenotypic values. Usually, a minimum number of 100 individuals is enough to have enough resolution on GWAS (Kumar et al., 2012), better if they represent a wide genetic diversity and are genetically distant. However strong GWAS analysis may consider using panels of at least 300-400 individuals up to thousands.
An important parameters to considered in GWAS analysis is the population structure, the relatedness between individuals need to be taken into account in order to avoid false positive

results due to the fact that not all individuals are not unrelated (Alqudah et al., 2020). Rather, a residual population genetic structure is always present in panels/germplasm collections and therefore this can generate spurious marker-trait associations or false positives.
The population structure analysis can be performed with population structure software such as STRUCTURE (Pritchard et al., 2000) and ADMIXTURE (Alexander et al., 2009), or by computing principal component analysis (PCA). The population structure analysis via STRUCTURE or ADMIXTURE softwares allowas to define a Q matrix where the number of Qs represents the different subpopulations detected in the panel, each accession has a value which estimates the proportion of membership for each subpopulation (Pritchard et al., 2000). The PCA is computed with the eigenstrat method, where genotypes are clustered based on the variance components where the major part of total variation is represented between PCA1 and PCA2 (Price et al., 2006). Another method is to calculate the genetic similarity between individuals calculating the relationships using genotypic data. This is performed via a mixed model approach where high values of relatedness correspond to high genetic similarity, reported in a kinship matrix (K) (51). The GWAS analysis, to correct for population structure, usually uses a combination of population structures (Q) and kinship matrix (K)(Bayer et al., 2017; Cockram et al., 2010).
Another very important factor to be considered is the allele frequency, as rare alleles are responsible for GWAS lack of resolutions. Rare allele are defined as allelic variation with a frequency lower than 5% in the population, they can explain favourable genetic variance of a small group of individuals in the panel (Alqudah et al., 2020). Usually, the GWAS analysis filter alleles for a minor allele frequency (MAF) major of 5%, or by a minimum number of individuals, loosing the power to identify rare alleles. Thus, a possible solution could be to decrease the MAF filtering or to increase the number of individuals, introducing wild relatives to increase the number of SNPs and resolution power of GWAS analysis (Alqudah et al., 2020).
As the GWAS analysis deals with genotypes that can be genetically distant, the LD decay between the accessions has to be estimated. The LD indicates both the interval of associated SNPs to restrict the QTL confidence interval and to estimate the distance between molecular markers (SNP) which is used to understand the number of molecular markers to be used (Myles et al., 2009; “The genetic dissection of quantitative traits in crops | Semagn | Electronic Journal of Biotechnology,” n.d.). The association score is usually measured with r^2 value, this parameter is useful to estimate both which molecular markers are inherited together and to retain markers with r^2 value higher than 0.2 or 0.3 in GWAS analysis to detect QTLs (Alqudah et al., 2020).
Both population structure and linkage disequilibrium decay rate are population-specific parameters. Therefore, each population should be preliminary assessed for population structure and LD-decay rate.
All these parameters are considered in GWAS analysis using different analysis tools, such as GAPIT3 (Wang and Zhang, 2021) and TASSEL5 (Bradbury et al., 2007). For example, GAPIT3 allows to merge results from phenotypic analysis (BLUEs or BLUPs) and genotypic data to perform GWAS analysis and detect associated QTLs using different statistical models (Wang and Zhang, 2021; Alqudah et al., 2020). Among the different models that could be used, GLM, MLM, MLMM, FarmCPU and BLINK are worth pointing out. The general linear model (GLM) does not use population structure and considers all the individuals as one group. The mixed linear model (MLM) uses fixed effect and individuals as random effects, taking in consideration population structure as covariate based on K or Q matrixes.
The multiple locus mixed linear model (MLMM) use forward-backward stepwise linear mixed-model regression to include associated markers as covariates to search for other markers masked from the associated marker effect. The Fixed and random model Circulating Probability Unification (FarmCPU) has a higher statistical power than MLMM and does not use kinship or Q matrixes to

avoid false positive markers. In fact, the markers detected by the iterations are fitted as cofactors to control false positives for testing the rest markers in a fixed effect model. To avoid the over model fitting problem in stepwise regression, a random effect model is used to select the associated markers using maximum likelihood method (Liu et al., 2016). An implementation of FarmCPU is represented by BLINK which has an augmented statistical power and computational efficiency. BLINK, in comparison to FarmCPU, uses Bayesian Information Content (BIC) of a fixed effect model to approximate the maximum likelihood of a random effect model to select the associated markers (Huang et al., 2019).

GWAS analysis establish a significant threshold for the -logPvalue or logarithm of odds (LOD), which corresponds to the logarithm between the ratio of the probability that the marker is significantly associated with the phenotypic variation on the probability that the association casually occurred (H_0). The usual threshold value is 3, meaning that the probability of significant association is 1000 times higher than the null hypothesis. This value is usually calculated with the Bonferroni correction formula, which is the logarithm of the ratio between the significant p-value threshold (0.05) for the number of non-redundant markers at each locus (Maurer et al., 2016; Saade et al., 2017).

The output results from GAPIT are represented by the Manhattan plots, which report the -log(p-values) of all markers divided by chromosomes and the threshold calculated by Bonferroni adjusted formula. Other important outputs are the QQ plots, which reports the deviations between the expected and registered pvalues for each markers, showing how all the models sufficiently correct for covariates and population structure avoiding false positive peaks (Alqudah et al., 2020).

After the definition of major QTL by GWAS analysis, further analysis could be performed in order to fine map the QTL confidence interval and MAS on selected panel based on the development of informative molecular markers for the trait of interest. Nowadays, among the highly diffused molecular markers, Kompetitive Allele-Specific PCR (KASP) genotyping is of major importance for marker assisted selection and QTL representing and high efficient and cost effective technology (Woodward, 2014). The KASP technology is based on three primers, two of them are allelic specific designed on the informative SNP, and the third one is the common primer genomic specific. The SNP specific primers work alternatively with the common primer, depending on the specific allele, and have a FRET cassette at 5’ end with different fluorophores such as FAM, HEX, VIC, TET, JOE, ROX, PET (Kaur et al., 2020). The KASP genotyping analysis is performed by endpoint PCR where both alleles of target SNP can be detected in a single reaction, first introduced by LGC Genomics (Semagn et al., 2014). This technology is more efficient and more advantageous in comparison to other precedent molecular markers such as SSR, RAPD, STS, AFLP because they are more time consuming and require additional molecular techniques to be analysed (such as agarose or acrylamide gels) (Neelam et al., 2013).

3. Aim of the thesis

The aim of the thesis is principally focused on the H2020 Innovar project "Next generation variety testing for improved cropping on European farmland" (2019-2024). The Innovar project is a European project which involves different research institutes and breeding companies different Europe. The principal aim of the project is to innovate the procedure of wheat varietal registration using genomics, phenomics and machine learning technologies. Briefly, European Community of Plant Variety Office (CPVO) commission has established protocols to describe varietal registration procedure for durum wheat and other several crops, commercially distributed and with a relevant value in the European market. As regards wheat, a variety can be registered as new if it demonstrates its distinctness, uniformity and stability (DUS) criteria and its value of cultivation and use (VCU). The VCU and DUS are needed for a variety to be subscribed in the national and European registers, thus being commercialized in Europe. In addition, a breeder can decide to protect the variety by demonstrating its distinctiness (DUS) and novelty, acquiring plant variety protection (PVP) and plant breeders right (PBR). This is a separate procedure in comparison to the European commercialization, as the breeder can decide only to protect its variety (es. with CPVO regarding European territory) without commercializing it.

The DUS protocols for wheat rely on different morphological criteria about plant physiology and development, the VCU protocols are mainly related on yield traits. These protocols are tested in Innovar durum and bread wheat panels for 3 years field trials where the VCU and DUS phenotypic traits are obtained. The Innovar project, together with phenotypic data and phenomics UAV technology, exploits genotyping platforms such as Illumina iSelect Infinium SNP 90K Chip array to genotype the bread and durum wheat varieties. Genome wide association analysis (GWAS) is performed to detect highly informative SNP molecular markers, associated with major QTLs responsible for the phenotypic variance of each DUS and VCU trait. The phenotyping, phenomics and genomics data will be used by machine learning technology to perform genomics prediction and establish the phenotypic characteristics of varieties based on the presence of informative alleles in the identified major QTLs for DUS and VCU traits. At the end of the project, the molecular and genetics information obtained will be proposed to augment the currently used European DUS and VCU protocols with molecular markers and information directly connected with the agronomic traits included in the protocol. This project will facilitate the work by breeders for the variety's registration procedure, as it will deliver molecular markers that could be used in MAS to better discriminate agronomic traits. Furthermore, the phenomics and genomic prediction analysis will better characterize varieties performances, paving the way to innovative agriculture procedures which improve environmental sustainability and crop production. The acquired knowledges in the Innovar project, will be also transferred to other crops included the DUS and VCU protocol. The herein reported PhD thesis, is arranged in three chapters connected with Innovar objectives.

1. The first part deals with the first year of trials (2020/2021) for Innovar regarding the VCU and DUS durum panels. The aim of the project was to fully characterize the Innovar durum panel obtaining phenotypic data from the VCU and DUS protocols, and genotypic data exploiting Global durum resources in Unibo and the Illumina iSelect Infinium SNP 90K array (https://wheat.pw.usda.gov/GG3/global_durum_genomic_resources). The acquired data were used to perform GWAS analysis and the final result was the identification of different major QTLs and candidate genes involved in yield related traits for VCU trials and in morphological traits for DUS trials.
2. As previously mentioned, the VCU protocol is mainly focused on yield related traits, but the evaluation of wheat diseases is also considered. However, the second main part of the PhD

thesis was to evaluate the Innovar durum panel and related mapping populations for the resistance to Soil Borne Cereal Mosaic Virus (SBCMV), a viral disease which causes severe yield losses in Northern Italy and Europe. The aim was to identify candidate genes involved in the resistance reaction starting from the cloned QTL *sbm2* responsible for the resistance, and to evaluate the presence of resistant haplotypes in Innovar durum panel.

3. As regards morphology traits, some Innovar durum varieties were also evaluated for spike fertility trait controlled by the QTL *GNI-2A,* responsible for the increase in the number of fertile florets per central spikelets. The objective of the third chapter was to identify the *GNI-2A* haplotype in Unibo durum panel, which includes a major amount of Innovar varieties, and on a segregant biparental population (Relief x Iride) that are connected with spike architecture and yield increase. The final outcome was the development of molecular markers connected with fertility trait that can be used to discriminate varieties haplotypes for DUS field trials. In fact, different phenotypic traits monitored in DUS protocols are connected with spike morphology and development. By designing KASP markers connected with the fertility haplotype determined by *GNI-2A* QTL, molecular markers can be connected to spike development DUS traits and can be used to better select and discriminate best performing varieties based on spike fertility and architecture.

4. GWAS analysis on durum panel for VCU and DUS agronomic trait characterization in Innovar durum panel

4.1 Introduction

4.1.1 Wheat variety registration procedures in EU – DUS and VCU protocols

Innovation in agriculture and the development of plant breeding increased during the nineteenth century, thus increasing the number of commercial varieties in the market and guaranteeing the growing population supply of food (Jamali et al., 2020; Bharadwaj et al., 2016). The increase in agricultural production was a result also of the innovation in technologies, use of fertilizers and pesticide products. The genetic improvement objectives and methodologies has to necessarily evolve and change together with the advancements in agriculture management techniques.
However, intensive agriculture has led also to unfavourable effects due to the fossil fuels, groundwater contamination and toxic components of chemicals. In addition, climate change hampers the agricultural system with drought seasons, high temperature and excessive rainfalls (Bharadwaj et al., 2016). Recent reports show that there will be an increase in population of around 10 billion by 2050, following the need of meeting global food demands and fuel request, meaning that there is an urgent need of increase production (Godfray et al., 2010). Food security should be obtained by increasing production to almost 80% by 2050, putting a lot of pressure on breeding programs that should be incremented 2.5 times respect to the current rate (Noleppa and Cartsburg, 2021; Stamp and Visser, 2012). In particular, cereals represent a staple cultivar for world agriculture and population diet, representing 56% of the world's calories as food and 44% for animal feed (http://www.fao.org/DOCREP/006/Y4683E/y4683e06.htm#TopOfPage).
Given this trend regarding environmental data and food supplies, breeders are subjected to a strong pressing to develop new varieties capable of satisfying agricultural needs. The aim is to register new breeding varieties resilient to biotic and abiotic stresses, and with increase yield and production. Generally speaking, current varieties must adapt to environmental changes and, at the same time, producing more yield per hectare and being resistant to pests and diseases in addition to drought (Bharadwaj et al., 2016; van Elsen et al., 2013; Miflin, n.d.). Plant breeders have been focusing on selecting varieties with economical income, resistant to biotic and abiotic stresses, with the desired technological quality and with high yield potential (van Elsen et al., 2013).
As regards to abiotic stresses, such as heat resistance and drought tolerance, genetic improvement and breeding programs plays an important role in selecting resilient varieties to adapt to the current climate change (Bharadwaj et al., 2016; Miflin, n.d.). In addition, also the selection and registration of varieties resistant to abiotic stresses has an important role in stabilizing yield component in plants (van Elsen et al., 2013).
As regards to yield, it is directly involved in food security and qualitative plant features. The registration of more stable varieties positively affects yield, which is increased also by the selection of varieties resistant to biotic stresses such as viruses, fungus, nematodes and bacteria (Elsen et al., 2013). Beside food security, plant breeding also focuses on food safety by reducing possibilities, for plant varieties, to the uptake of heavy metals but also mycotoxins, especially in cereals (Redman and Noleppa, 2017).

To conclude, the mentioned traits are some of the agronomic characteristics that are pursued by breeders and by EU protocols to produce innovative varieties that face climate change, innovation in technology and agricultural market objectives.

In order to select new varieties, breeders perform artificial crosses between two or more different parental lines with favourable agronomic traits which are known to be stable for multiple years and different environments (Yu and Chung, 2021). Typically, several hundreds of crosses are carried out per year in a breeding company. During the selection steps for different generations, the genotypes with undesired characteristics are discarded, but different genetic combination could be generated identifying new varieties ("Principles of Plant Genetics and Breeding, 2nd Edition | Wiley," n.d.). This process is time and effort consuming and requires approximately ten years to register a new variety (Jamali et al., 2019; "Principles of Plant Genetics and Breeding, 2nd Edition | Wiley," n.d.). In order to protect the new breeding varieties and their materials, plant breeders have different possibilities to register their varieties and acquire royalties (Butruille et al., 2015; da Silva et al., 2017; Glenn et al., 2017). In order to guarantee the plant variety protections, two organisms are involved: the Community of Plant Variety Office (CPVO), a decentralized EU agency instituted in 1995, and the International Union for the Protection of new Varieties (UPOV). The CPVO represents the EU protection system which regulates and controls the plant breeders right (PBRs). The UPOV is an international organization composed of 78 different countries, which are responsible for stabilizing the principles and guidelines (UPOV act 1991) to guarantee PBR and PVP to newly bred varieties (Cooke and Reeves, 2003).

The UPOV establish the Plant Variety Protection system (PVP), with the aim to organize the Intellectual property rights (IPR) among the member states. As regards to variety registration, the UPOV act of 1991, recognizes the plant breeders right (PBR), which are valid for 20 years based on the species (for example, tree crops have a longer breeding protection), and the plant variety rights (PVR). Furthermore, the European Union protects the IPR by the Community Plant Variety Rights system (CPVR), coexisting with the PBR system (UPOV act, 1991).

The whole system is managed by the CPVO, a decentralized EU agency instituted in 1995.

The different steps and regulation for variety registration protocols are shared among European countries, linked both to variety development, to the seed certification and to final distribution ("'Framework for the introduction of Plant Breeder's Rights, Guidance for practical implementation' | Naktuinbouw," n.d.). The EU presents the most wide protocols for variety registration, which extend to different crops, vegetables and fruit species (Jamali et al., 2019). The EU system for variety registration is based on models for novelty and value of varieties, thus, if national variety list fulfil these requirements, it is directly included on the EU common catalogue (Gilliland and Gensollen, 2010) .

The UPOV established three criteria to recognize the PVP to a new variety. In fact, the new variety must be distinct from the others, homogeneous and remain stable over multiple generations (Jördens, 2005). Breeders can apply to PVP as long as they follow the UPOV criteria and do not perform one of the following activities without other's breeder permission: sale of a variety protected by PVP, multiplication of the variety without permission, hybrid production using other varieties and distribute PVP varieties without permission (Yu and Chung, 2021).

However, there are some exemptions that occur for breeders, for example, varieties could be used as source of initial variation, for research purposes and for private use (Jördens, 2005).

In order to be awarded by PVP, breeders need to select varieties which follow some key principles determined by UPOV convention: distinctiveness (D) from other varieties, uniformity (U) in its agronomic characters between individuals reported in the descriptive report of the specific varieties (included in the CPVO protocol) and stability (S) over different generations (DUS). This principles are

established by the UPOV through the DUS phenotypic test, which are mandatory for breeders in order to register a new variety (Glenn et al., 2017; Jördens, 2005; Yu and Chung, 2021).
The DUS trials could be performed either in greenhouse or field, for two years or at least in different locations, where different morphological characters are monitored to verify the presence of enough diversity between the new varieties to be certified and the most similar ones (Yu and Chung, 2021). The list of morphological characters to monitor is described in the UPOV protocols with detailed guidelines, different for every plant species (Jördens, 2005). For example, considering species such as wheat and barley, varieties need to be vernalized before performing the DUS trials (Yu and Chung, 2021).
The evaluation of the DUS characters relies on different examination, namely formal, substantive and technical (European Union Intellectual Property Office and Community Plant Variety Office., 2022). The first two type of examinations are carried out by the CPVO, concerning formal and fulfilment of specific conditions. The technical examination is performed by CPVO delegate offices and is aimed at evaluating that the requirements for DUS variety registration are fulfilled. In case of positive outcome after these examinations, the official description of the variety is carried out and registration occurs (Cooke and Reeves, 2003; European Union Intellectual Property Office. and Community Plant Variety Office., 2022).
The DUS phenotypic trials have some limitations in relation to the morphological characters that are monitored. In fact, the measured phenotypic traits could be influenced by environmental conditions which could hamper the evaluation and expression of the phenotypic characters (Cooke and Reeves, 2003; Wang et al., 2016). Moreover, the DUS tests do not consider the actual agronomic value of a cultivar, and only in part the DUS allocate traits / characters of agronomic relevance.
Different studies were published regarding the use of molecular markers to discriminate morphological characters. In particular, it has been shown how SSR markers could be precisely used to distinguish between different DUS traits, or between genotypes with different pedigree in a consistent way in comparison to morphological trait evaluation (Yu and Chung, 2021). As regards the integrated use of phenotypic traits and molecular markers to discriminate varieties as distinct in DUS protocols, varieties with molecular markers similarity higher than 96% but different for morphological states are considered as distinct in order to not increase the need of further resources during breeding processes. However, very few markers have currently been proposed for DUS discrimination (Yang et al., 2021). Beside the use of single molecular markers like SSR, different studies also proved that the outbreak in NGS technologies, SNP arrays and genome editing are being widely used by breeding programs for marker assisted selections (MAS) and variety selection (Achard et al., 2020; Cockram et al., 2012; Sarao et al., 2010; Song et al., 2003; Tian et al., 2015). However, the UPOV doesn't recognize the use of molecular markers as a principal source to use for variety discrimination, exploiting them just in case they correlate perfectly with DUS characters, strongly decreasing their advantage in genotype profiling capacity (Yu and Chung, 2021). This was strengthen by Yang et al (Yang et al., 2022), who reported that the DUS system is very variable and lacking in consistency across environments suggesting the increase in genomic data for DUS trials.
Following the DUS trials, other criteria necessary in the EU to register a variety are described as Value of Cultivation and Use (VCU), however the international coordination is not as precise as it is for DUS trials (Cooke and Reeves, 2003). The VCU is not mandatory for a variety to be recognized with PVP, but it is required for that variety to be included in the National/European catalogue and to enter in the market. Basically, a variety must be recognized as carrier of improved qualities in its value of cultivation or use, no matter in which environment the VCU field trials are performed knowing that the EU requires different trials and replicated fields (Cooke and Reeves, 2003). As regards VCU, different agronomic traits are considered for the major part on yield components and disease resistance traits and quality parameters. VCU tests need to evaluate the utility of the variety

also from the genetic point of view, representing an important resource for the marker and policymakers.
Data obtained from VCU and DUS all together enable breeders and examination organs to discriminate between different varieties, describing each accessions for any particular morphological and marketing improvement in comparison to other already registered varieties (Cooke and Reeves, 2003).
In Italy, the registers containing all the information on the varieties can be consulted in the database of the National Agricultural Information System (SIAN) and are updated periodically. Currently the number of total varieties entered in the registers is 6146 of which 43% is represented by cereals.
In particular, durum wheat (*Triticum turgidum* L. ssp. Durum) is the most developed species in Italy, with about 4 million tons on average produced each year, mainly used in pasta production (IWGSC, 2014). In Italy, several companies at national level operate in the seed sector for the genetic improvement of wheat, selecting the best varieties. Until now, 290 varieties of durum wheat (http://www.sementi.it/) are certified in the national registers.

4.2 Materials and methods

4.2.1 Germplasm collection

The Innovar germplasm durum collection was composed of modern durum wheat varieties, coming from different breeding programs and commercially available in the market. Varieties were chosen based on the optimal environmental adaptation capacity of each variety, on important characteristics in relation to disease resistance and interesting agronomic traits and on the high diffusion and utilization in the seed market.
Different varieties were chosen based on the different durum trial locations of Innovar project and trial layouts. The Innovar trials were divided in value for cultivation and use (VCU) trials and in distinctiveness, uniformity and stability (DUS) trials, tested across different locations. The VCU and DUS trials were phenotyped following specific protocols. As regards the DUS, the protocol followed CPVO European directives. On the other hand, the VCU protocol was defined by Innovar partners after consultation of different protocols adopted by few European nations.
Germplasm seeds collected for genotyping were sown in Innovar durum wheat VCU trials, which took place in different agroclimatic zones (ACZ) based on previous climate studies (Ceglar et al., 2019; Trnka et al., 2011). The VCU trials were divided between core trials and drought trials across different locations. The durum wheat trial sites, as regards VCU core trial, were two for Maritime South ACZ (Horta and UNITUS, Italy), one for Mediterranean ACZ (UPM, Spain) and one for Pannonian ACZ (UNIDEB, Hungary). Furthermore, as regards VCU drought trials, one location was in Maritime South ACZ (UNITUS, Italy), two in Mediterranean ACZ (CSIC, ICARDA, and UNITUS, in Spain and Morocco and Italy respectively) (Figure 2 a and b).

a) b)

Figure 2: durum wheat ACZ in INNOVAR trials. a) Mediterranean (red), Maritime South (yellow), Pannonian (green). b) Mediterranean (red), Maritime South (yellow).

Regarding durum wheat, 166 different varieties were sown in 3 ACZ both in VCU core and VCU drought trials: Maritime South, Mediterranean, and Pannonian. In durum wheat trials, 5 varieties were common between all ACZ, 5 within each ACZ, and 35 specifics for each trial site (Table S13). However, different varieties were shared between locations, especially between Spain and Italy which share similar climate conditions.

In addition to VCU trials, DUS trial was carried out in collaboration with CREA-DC, in the field station of Tavazzano (Italy). The total number of VCU varieties was included in the DUS trial, with the addition of further durum germplasm varieties from Unibo background material to augment genetic diversity (Table S14). Unibo varieties were included as important source of genetic diversity and carrier of relevant agronomic characteristics. Furthermore, they were mainly included from the durum panel 1 (DP1) (Maccaferri et al., 2015) which was also characterized in different project both phenotypically and genotypically with Illumina iSelect Infinium 90K SNP Chip array.

In addition, the DUS trial was augmented with CPVO example varieties and others from CREA-DC collection. The CPVO varieties were directly selected from CPVO protocols in order to have specific controls to compare for the different agronomic trait's evaluation in DUS tests (Table 1). The final number of varieties tested in the DUS field trial was 253.

Table 1: list of CPVO contriols and CREA collection varieties included in the DUS panel, tested in CREA-DC.

Varieties	Additional types
Amilcar	CPVO controls
Arcangelo	CPVO controls
Atoudur	CPVO controls
Auradur	CPVO controls
Bolo	CPVO controls
Carioca	CPVO controls
Carpio	CPVO controls
Ciccio	CPVO controls
Colosseo	CPVO controls
Creso	CPVO controls
Don_Ricardo	CPVO controls
Duilio	CPVO controls
Elsadur	CPVO controls
Iride	CPVO controls
Karur	CPVO controls
Kiko_Nick	CPVO controls
Levante	CPVO controls
Lupidur	CPVO controls
Meridiano	CPVO controls
Orobel	CPVO controls
Simeto	CPVO controls
SY_Lido	CPVO controls
Tiziana	CPVO controls
Aceres	CPVO controls
Arcobaleno	CPVO controls
Asdrubal	CPVO controls
Cantico	CPVO controls
Canyon	CPVO controls
Chiara	CPVO controls
Grecale	CPVO controls
Italo	CPVO controls
Ofanto	CPVO controls
Verace	CREA Collection
Matusalem	CREA Collection
Neruda	CREA Collection
Sorrento	CREA Collection
SY_Esperto	CREA Collection
Baronio	CREA Collection
RGT_Natur	CREA Collection
Ottaviano	CREA Collection
Solstizio	CREA Collection
Brancaleone	CREA Collection
Dario	CREA Collection
Magellano	CREA Collection

To summarize, phenotypic and genotypic data were collected from the environments and number of varieties reported in each trial, reaching a final number of 166 unique varieties for VCU durum dataset and 253 varieties for DUS durum dataset (Table 2).

Table 2: VCU and DUS environments and specific varieties for each trial. The different field trials are grouped in Innovar agro-climatic zones (ACZ).

VCU trials	Agro Climatic Zones (ACZ)	Field trial	Number of genotype	Total unique genotypes
VCU Core trials	**Maritime South**	Horta (Foggia, Italy)	45	166
	Maritime South	UNITUS (Viterbo, Italy)	45	
	Mediterranean	UPM (Escacena, Spain)	45	
	Pannonian	UNIDEB (Nyíregyháza, Hungary)	45	
VCU Drought trials	**Mediterranean**	CSIC (Santaella, Spain)	30	
	Maritime South	UNITUS (Viterbo, Italy)	30	
DUS trial	**Maritime South**	CREA-DC (Tavazzano, Italy)	253	253

4.2.2 VCU trial layout and phenotyping

The VCU trial layout was composed, for each location, with a split plot design. Each trial was characterized by two fungicide treatments, with three replicates each and complete randomized varieties within each replicate. As regards the VCU core trials, the main objective was to assess the different varieties in each trial for yield performances and disease resistance, parameters evaluated in the VCU protocols, under two levels of fungicide regime. The two levels were composed of fungicide (Prosaro X wide spectrum fungicide, different depending on locations and on common agricultural practices) used as full (complete application) and minimal (no application) treatments.
In details, in the full treatment, varieties were expected to be evaluated for full yield potential with no yield loss due to disease, straw breakdown or pest damage. On the other hand, the minimal treatment received no fungicide, thus allowing individual varieties' susceptibility to disease and straw damage to be monitored. The difference in yield between full and minimal treatments will provide a calculation of risk of yield loss due to disease and straw damage.
The plot size across different locations derived from 10 m^2 to 15 m^2 and the used seed density followed local practices (generally 400 seeds/m^2).

Across all the locations, the trial layout was similar to the one reported in (Figure 3), where the VCU core trial in HORTA (Foggia, Italy) field station was taken as example.

										Replicate	Treatments
70	90	89	73	66	71	86	69	88	56	R3	Full
53	59	54	92	72	98	52	82	51	61	R3	Full
83	93	75	63	78	62	64	99	67	55	R3	Full
74	79	94	97	96	65	85	100	57	58	R3	Full
68	95	81	91	76	77	80	60	84	87	R3	Full
15	48	19	37	21	38	20	24	22	41	R3	Minimal
10	5	8	32	7	4	3	36	27	28	R3	Minimal
35	1	49	11	18	23	29	6	26	16	R3	Minimal
30	25	45	12	17	39	9	44	31	33	R3	Minimal
14	34	50	2	42	43	40	47	13	46	R3	Minimal
37	39	26	46	30	2	25	31	5	21	R2	Minimal
47	6	10	18	49	41	42	15	35	3	R2	Minimal
19	8	48	36	50	1	11	34	12	43	R2	Minimal
38	17	7	16	9	24	13	33	32	4	R2	Minimal
27	28	40	20	29	44	23	14	45	22	R2	Minimal
59	80	77	88	98	64	55	91	93	85	R2	Full
76	86	57	78	89	84	81	61	75	65	R2	Full
67	66	87	90	60	72	54	51	73	94	R2	Full
79	96	68	58	69	82	71	62	92	74	R2	Full
99	70	56	100	97	53	83	63	52	95	R2	Full
91	92	93	94	95	96	97	98	99	100	R1	Full
81	82	83	84	85	86	87	88	89	90	R1	Full
71	72	73	74	75	76	77	78	79	80	R1	Full
61	62	63	64	65	66	67	68	69	70	R1	Full
51	52	53	54	55	56	57	58	59	60	R1	Full
41	42	43	44	45	46	47	48	49	50	R1	Minimal
31	32	33	34	35	36	37	38	39	40	R1	Minimal
21	22	23	24	25	26	27	28	29	30	R1	Minimal
11	12	13	14	15	16	17	18	19	20	R1	Minimal
1	2	3	4	5	6	7	8	9	10	R1	Minimal

Figure 3: field trial design of Horta-Foggia VCU durum core trial. Varieties were completely randomized within each replicate (3), under different treatments (full=green and minimal=yellow). Then number of varieties is reported in each cell.

As regards the VCU drought trials, the number of replicates and plots size was maintained equal to VCU core trial. On the other hand, the fungicide was maintained as “full treatment”, but two drought regimes were performed in three replicates each (using the same varieties sown for the VCU core trial): full irrigation and not irrigated treatments.

The VCU phenotypic protocol was equal both for VCU core and VCU drought trial, monitoring different traits at specific growth stages following Zadoks BBCH scale (Zadoks et al., 1974) (Figure 4 and Table 3)

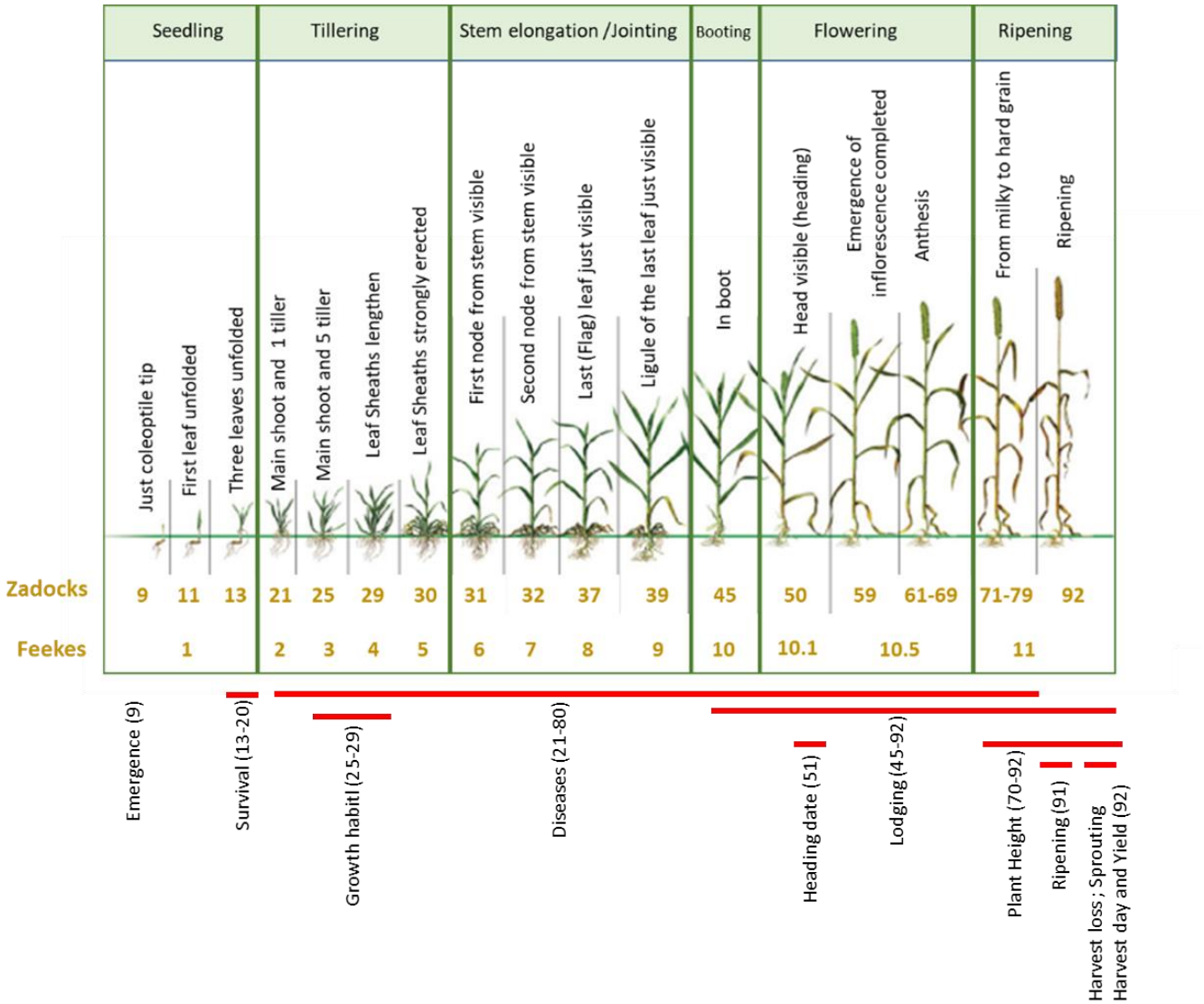


Figure 4: Zadoks BBCH scale of wheat. The phenotypic measurements for VCU trials and respective growth stages are reported with red lines.

Table 3: list of specific measurements and layout characteristics collected for VCU trials across the different environments at the different growth stages (from 10 to 93).

	Character	Zadocks	Innovar protocol
Trial Establishment	**Design**	-	Factorial, split-plot design
	Replications	-	3
	Plot Size	-	Minimum 10 m^2 Optimun 15 m^2
	Distance between rows	-	According to local practices
	Seed density	-	According to local practices
Crop development	**Sowing date**	-	day/month/year
	Emergence	9-11	day/month/year (date of 75% of plots with 50% emergence)
	Survival or Plant population	13-20	1-9 scale (9= total survival)
	Growth habit	25-29	1-9 scale (9= postrate)
	Heading Date	51	Days (days from sowing at which 50% plants in the plots have first spikelet visible on ears)
	Plant Height (at harvesting)	70-92	From ground level to the top of the ear/panicles, ignoring awns.
	Ripening	91	When the earliest varity reach GS91, record the GS of each plot
Incidences	**Lodging/Leaning** (at harvesting)	45-92	% of the plot
	Winter hardiness/damage	any time	%
	Harvest Loss	92	Number of grains in $1m^2$
	Sprouting	92	%
	Bird/other damages	any time	% of the plot
Diseases	**Yellow Rust**	21-80	% infection
	Brown rust	21-80	% infection
	Septoria	21-80	% infection
	Fusarium ear blight	21-80	% infection
	Powdery mildew	21-80	% infection
Harvest	**Harvest date**	92	day/month/year
	Yield	92	kg/ha with 2 decimals
	Moisture content	-	%
	Specific weight	-	kg/ hl
	Thousand Grain Weight	-	g

Table 3 reports a summary of the different traits measured in VCU trials across all the environments and ACZ, with the specific scale of measurement and the general features of the trial layouts. Basically, presence of diseases was monitored during the growing season at different stages, if different damages and other agronomic traits, such as crop development, incidences and, at the end, harvest traits.
The final yield was normalized to the 15% of moisture content with the following formula:

Yield_15 = Yield*(100-MoustureContent) / (100-15)

4.2.3 DUS trial layout and phenotyping

The DUS trial was carried out in collaboration with CREA-DC in the field station of Tavazzano (Italy). The selected varieties from the different VCU field trials were all included in the DUS, with the addition of background material from durum panel Unibo (DP1) (Maccaferri et al., 2015), CPVO controls and varieties from CREA-DC collection, in order to reach the final number of 253 genotypes (Table S14).
The varieties were tested in two replicates with alpha-design trial layout (Figure 5), dividing the field in blocks inside each replicate.

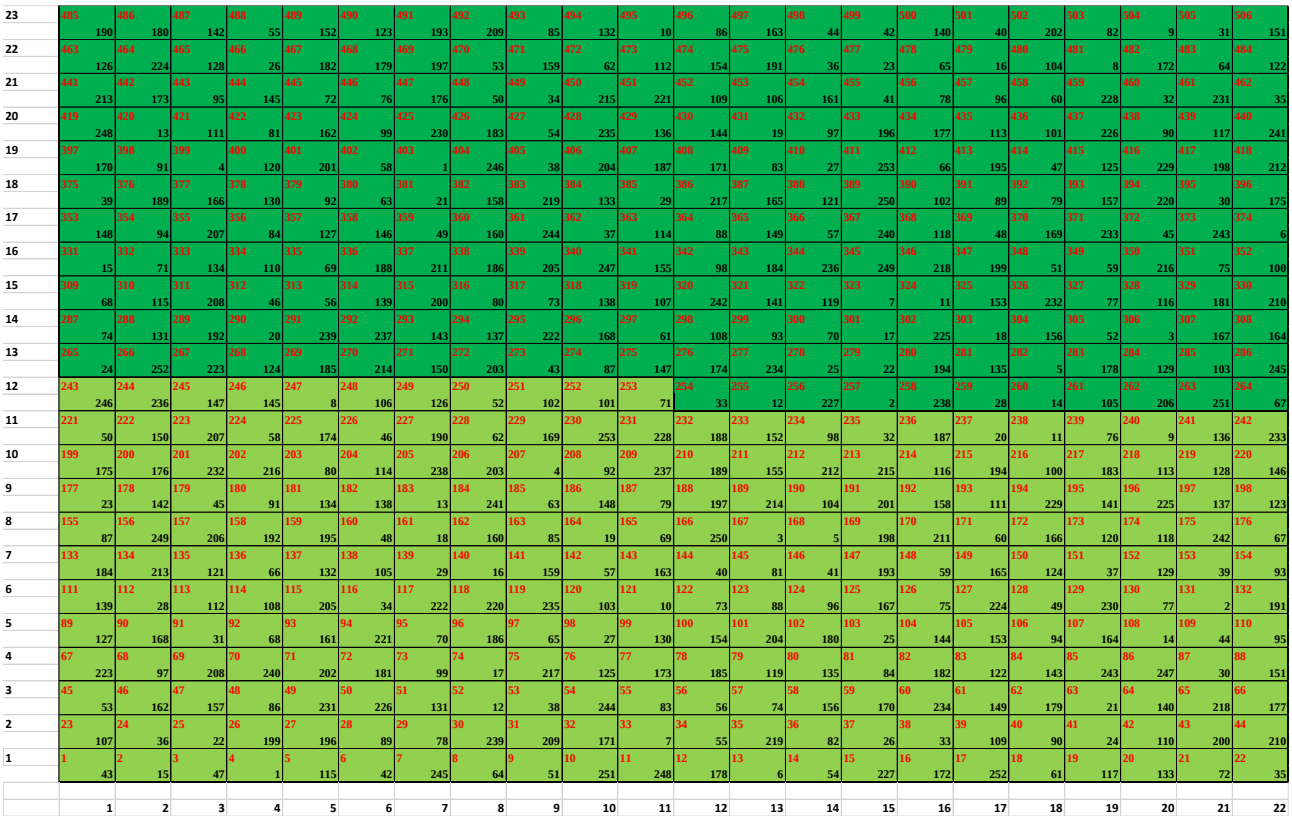


Figure 5: trial layout of DUS in CREA-DC with alpha design trial scheme. Green and light green were used to show the different replicates used in the trial. Red numbers correspond to the progressive plot number. Black number correspond to the genotype number.

The phenotypic evaluation occurred based on the CPVO DUS protocol at specific growth stages, evaluating each variety for the different morphological traits reported in the European CPVO protocol CPVO-TP/120/3 of 2014 (Table 4).
In details, most phenotypic traits were scored in the field trial with some exceptions regarding the seed and coleoptile colours, that were analysed in the lab with standard protocols.

Table 4: EU CPVO protocol for DUS trials. Measured traits were further characterized as follows: QL= Qualitative characteristic, QN = Quantitative characteristic, PQ = Pseudo-qualitative characteristic, MG = Single measurement of a group of plants or parts of plants for the assessment of distinctness, MS = Measurement of a number of individual plants or parts of plants for the assessment of distinctness, VG = Visual assessment by a single observation of a group of plants or parts of plants for the assessment of distinctness, VS = Visual assessment by observation of individual plants or parts of plants for the assessment of distinctness, A = Sample size of 100 plants to be observed for the assessment of uniformity, B = Sample size of at least 2000 plants which should be divided between at least two replicates, C = The assessment of the characteristic "Seasonal type" should be carried out on at least 300 plants.

Trait number	Growing stage	Trait	Range of traits	Type of characteristics	Type of assessment	Place of trait assessment
1	0	**Seed:** colour	1. white 2. reddish 3. purple 4. bluish	**PQ**	VG	Lab.
2	0	**Seed:** coloration with phenol	1. absent or very light 3. light 5. medium 7. dark 9. very dark	**QN**	VG	Lab.
3	0	**Coleoptile:** anthocyanin coloration	1. absent or very weak 3. weak 5. medium 7. strong 9. very strong	**QN**	VG	Lab.
4	25-29	**Plant:** growth habit	1. erect 3. semi erect 5. intermediate 7. semi prostrate 9. prostrate	**QN**	VG	Field
5	47-51	**Plant:** frequency of plants with recurved flag leaves	1. absent or very low 3. low 5. medium 7. high 9. very high	**QN**	VG	Field

6	49-60	**Flag leaf:** anthocyanin coloration of auricles	1. absent or very weak 2. medium 3. strong	**QN**	VG	Field
7.a	50-52	Time of ear emergence	Julian date (at least 3 times per week)		MG	Field
7 (calculated from 7a)	50-52	Time of ear emergence	1. very early 3. early 5. medium 7. late 9. very late	**QN**	MG	Field
8	60-65	**Flag leaf:** glaucosity of sheath	1. absent or very weak 3. weak 5. medium 7. strong 9. very strong	**QN**	VG	Field
9	60-65	**Flag leaf:** glaucosity of blade	1. absent or very weak 3. weak 5. medium 7. strong 9. very strong	**QN**	VG	Field
10	60-69	**Ear:** glaucosity	1. absent or very weak 3. weak 5. medium 7. strong 9. very strong	**QN**	VG	Field
11	60-69	**Culm:** glaucosity of neck	1. absent or very weak 3. weak 5. medium 7. strong 9. very strong	**QN**	VG	Field
12	69-92	**Lower glume:** hairiness on external surface	1. absent 9. present	**QL**	VG	Field
13a	75-92	**Plant:** length	from soil surface to tip of awns or scurs	**QN**	MG	Field
13 Calculated from 13a	75-92	**Plant:** length	1. very short 3. short 5. medium 7. long 9. very long	**QN**	MG	Field
14	80-92		1. thin	**QN**	VG	Field/Lab.

		Straw: pith in cross section	2. medium 3. thick or filled			
15	80-92	**Ear:** density	1. very lax 3. lax 5. medium 7. dense 9. very dense	**QN**	VG	Field
16a	80-92	**Ear:** length	From base to tip of ear	**QN**	MS	Lab.
16 calculated from 16a	80-92	**Ear:** length	1. very short 3. short 5. medium 7. long 9. very long	**QN**	MS	Lab.
17	80-92	**Ear:** scurs or awns	1. both absent 2. scurs present 3. awns present	**QN**	VG	Field
18	80-92	**Ear:** lenght of scurs or awns	1. very short 3. short 5. medium 7. long 9. very long	**QN**	VG	Field
19	80-92	**Ear:** colour	1. white 2. coloured	**QL**	VG	Field
20	80-92	**Ear:** shape in profile	1. tapering 2. parallel sided 3. slightly clavate 4. strongly clavate 5. fusiform	**PQ**	VG	Field
21	80-92	**Apical rachis segment:** area of hairiness on convex surface	1. absent or very small 3. small 5. medium 7. large 9. very large	**QN**	VG	Lab.
22	80-92	**Lower glume:** shoulder width	1. absent or very narrow 3. narrow 5. medium 7. broad 9. very broad	**QN**	VG	Lab.

23	80-92	**Lower glume:** shoulder shape	1. strongly sloping 3. slightly sloping 5. horizontal 7. slightly elevated 9. strongly elevated	**QN**	VG	Lab.
24	80-92	**Lower glume:** length of beak	1. very short 3. short 5. medium 7. long 9. very long	**QN**	VG	Lab.
25	80-92	**Lower glume:** beak shape	1. straight 3. slightly curved 5. moderately curved 7. strongly curved 9. geniculate	**QN**	VG	Lab.
26	80-92	**Lower glume:** area of hairiness on internal surface	1. very small 3. medium 5. very large	**QN**	VG	Lab.
27		Seasonal type	1. winter type 2. alternative type 3. spring type	**PQ**	VG	Field

4.2.4 Statistical analysis on phenotypic data

The statistical analysis was performed both on VCU and DUS field data using the softwares R and Rstudio (RStudio Team,2020). Outliers were removed using the interquartile rules: the interquartile range (IQR) of the data was multiplied by 1.5, outliers were defined as values 1.5 *IQR above the third quartiles and 1.5 *IQR below the third quartile. The heritability for each trait was calculated using the package *heritability* in Rstudio.

Best linear unbiased estimations (BLUEs) were obtained using the R package *lme4* by mixed model analysis (*lmer* function). The parameters introduced in the model were different based on VCU and DUS trials.

As regards VCU, three phenotypic datasets were generated across all the enviroments, as the same phenotypic evaluation occurred for all the different trials: VCU core, VCU drought and VCU core and drought merged together. The different environments included in the model are composed of two trials in Spain and Marocco (Mediterranean ACZ), two trials in Italy (Maritime South ACZ) and one trial in Hungary (Pannonian ACZ). The variables included in the model for BLUEs extraction across

all the VCU trials were the following, using genotype as fixed and other parameters as random variables:

Phenotype ~ Genotype + Replicate + Enviroment + Treatment + Genotype:Enviroment + Environment:Treatment + Treatment:Replicate

Based on the following model, interaction with environment (E), across genotype and environment (GxE), across treatments and environments/replicates were evaluated and considered during standardization of phenotypic values obtaining the BLUEs. The significant interactions between E and GxE were obtained via ANOVA analysis using the significant p-value threshold of 0.05.
As regards the DUS analysis, the trial data come only from one year and one location (CREA-DC) and the same analysis was performed with the following model to obtain the BLUEs, using genotypes as fixed effects and other parameters as random:

Phenotype ~ Genotype + Replicate + Blocks + Genotype:Blocks + Genotype:Replicate + Replicate:Blocks

ANOVA analysis was performed also on DUS trial dataset to obtain the significance of interaction between phenotypes and field varieties.

4.2.5 DNA extraction of Innovar DUS durum panel

In order to genotype the whole DUS durum panel with SNP molecular markers exploiting the Illumina iSelect Infinium SNP 90K Chip (Wang et al., 2014), DNA was extracted from all the Innovar DUS and VCU varieties augmented with 131 accessions from Unibo collection Durum-Panel-2 (DP2) (background materials, updated and augmented version of the durum panel 1) composed of modern durum wheat varieties, generating the final Innovar durum panel genotyping collection. As a result, the final durum wheat collection sent for Illumina iSelect Infinium array 90K SNP Chip genotyping consisted of 336 varieties. DNA was extracted following the CTAB extraction protocol (Doyle and Doyle, 1987) with few adaptations. Briefly, 100 mg of lyophilized leaf tissue were grinded in liquid nitrogen from the whole augmented panel, after being sown in greenhouse on alveolar pots and harvested after 10 days when the first leaf of each genotype reached a length of approximately 10cm. DNA was extracted with CTAB protocol, in comparison to the standard protocol the purification step with IAC-Chloroform was performed two times, in order to extract better quality DNA.
DNA concentration was checked via gel electrophoresis and biophotometer, retaining samples with high concentration and quality ratios. DNA was then diluted at 100ng/µl to be genotyped with Illumina iSelect Infinium SNP 90K Chip.

4.2.6 SNP genotyping with Illumina iSelect Infinium SNP 90K wheat array

DNA was shipped to North Central Small Grains Genotyping Laboratory USDA- ARS-ETSARC (Dr. Jason Fiedler, USDA ARS Cereal Crops Research Unit, Fargo, North Dakota) for genotyping based on Illumina iSelect Infinium SNP 90K Chip (Wang et al., 2014).
Innovar DUS and DP2 genotyping dataset was merged with genotyping data already available in Unibo to increase the number of SNPs to be analysed, namely: Global Durum Panel (GDP), 475 modern varieties included out of 1020 total accessions and Durum Panel UNIBO (DP1), 290 accessions (Maccaferri et al., 2019b, 2015; Mazzucotelli et al., 2020). The principal to augment the Innovar durum dataset with genotypic data from GDP and DP1 was to increase the genetic diversity and better perform the SNP call.
The merged Illumina raw data (Innovar DUS-DP1-DP2-GDP, a total of 1052 accessions) was analysed for SNP genotype calling before removing durum wheat GDP varieties not included in the GDP panel. The raw intensity data (.idat) was imported in GenomeStudio Module Polyploid Genotyping 2.0, and a custom SNP manifest file (Wheat90k-ConsAkhunovKSU-15033654-A) was used to perform the SNP call using Polygentrain algorithm with the following parameters: cluster distance set to 0.07 and INBRED option to separate A and B clusters.
A total of 81K SNPs were called and filtered for quality using Gentrain score parameter higher than 0.6, reducing the number of SNPs to 50,000. A further filtering step was performed to retain only the correctly physically and genetically mapped SNPs (single locus, Mendelian) based on the *Triticum turgidum cv* Svevo genome, as in the 90K consensus map (Maccaferri et al., 2019b, 2015). The final number of filtered SNPs used for the analysis was 24108, and a hapmap file was built on 1052 accessions with A/B allelic variants.

4.2.7 Imputation and LD decay

Polymorphic information (PIC) content was calculated for the merged dataset using the following formula (Serrote et al., 2020):

1 – (MAF)2 – ((1-MAF)2

PIC measures the ability of a marker to detect polymorphisms and therefore has enormous importance in selecting markers for genetic studies (Serrote et al., 2020).
An Unibo internally developed R script was used to filter the hapmap file based on the following parameters: Minor Allele Frequency (MAF) greater than 0.01, SNPs missing call greater than 0.5, samples with a missing rate above 0.5. After filtering, the final hapmap file included 21940 SNPs. The genotyping dataset was then imputed using Beagle v5.4 (Browning et al., 2021) to assign A/B variants to miss SNPs based on their position and closer SNPs (Beagle 5.4 uses a linkage disequilibrium-based algorithm).
The imputed vcf file was used to calculate the linkage disequilibrium (LD) decay in the durum germplasm using the software TASSEL5 (Bradbury et al., 2007). The LD decay was plotted using three linkage thresholds (r^2 equals 0.3, 0.5, and 0.8) using a sliding window equals to the average number of markers for each chromosome.

4.2.8 Pruning and Population structure analysis

PLINK software (Chang et al., 2015) was used for the pruning step, removing redundant SNPs in high LD (using r^2 thresholds 0.3, 0.5, and 0.8) and generating three output files for the different r^2 thresholds.
The pruned output files were used for population structure analysis using the model-based likelihood method ADMIXTURE (Alexander et al., 2009). The method was optimized using the block relaxation algorithm, the quasi-Newton convergence acceleration method, and q = 3 secants (Alexander et al., 2009), defining the sub-population memberships from k=2 to k=20. To detect the best number of subpopulations to be analysed, the cross-validated error rate, delta cv error, minimum group size, maximum admixed lines in a group, and admixed lines percentage were considered.
TASSEL5 was used to convert the imputed hapmap file into a distance matrix and thus obtain the kinship matrix by converting the values in genetic similarities. The distance matrix was calculated using identity by state method and obtaining genetic distances (GD). GDs were converted to genetic similarities (GS) to build the kinship matrix using the following formula: GS = 1 - GD. Heatmap and ward clustering (Ward.D2 algorithm) were performed on the kinship matrix using R packages pheatmap v1.0.12 and dendextend v1.15.2.
Neighbour Joining Tree (NJ) was computed with the R package adegenet v2.1.5.

4.2.9 GWAS analysis

GWAS analysis was performed using the R package GAPIT3 (Wang and Zhang, 2021) with few edits to the pipeline.
Basically, the final imputed hapmap was filtered in separate files for DUS and VCU varieties, and separate kinship and population structure analysis were performed for the two datasets. Separate analyses were also performed regarding SNP pruning with the previously mentioned r^2 threshold used for the whole genotypic datasets which included Innovar DUS panel, DP1, DP2 and GDP.
Both DUS and VCU hapmaps were analysed with 1000 permutation steps performed with FarmCPU algorithm, in order to detect the best fitting logarithm of odds (LOD) threshold of significance for the QTLs peaks. The permutation threshold was compared to the Bonferroni adjusted threshold, obtained by dividing the significant p-value of 0.05 with the number of pruned markers at a r^2 threshold of 0.8 and calculating the negative logarithm in base 10.
The GWAS analysis was performed with GAPIT3 using the following models: GLM, MLM, MLMM, FarmCPU and Blink. The edited GAPIT3 pipeline was divided in two step analysis with different population structure correcting methods: the first step included the kinship matrix derived from TASSEL5, the second included both the kinship matrix and the Q file derived from the population structure analysis with ADMIXTURE, selecting the correct number of Ks based on the cross-validated error rate, Delta cv error, minimum group size, maximum admixed lines in a group, and admixed lines percentage were considered.
The first analysis included the following edited models: GLM (naïve), MLM + K, MLMM + K, FarmCPU and Blink.
The second analysis included the following models, where Q file from ADMIXTURE analysis were used as covariates: GLM, MLM + Q+ K, MLMM + Q + K, FarmCPU + Q, Blink + Q.
As for both pipelines, the number of dimensions for PCA was set to 0 and model selection to false.
The final Manhattan plots and raw data were merged in unique file including all the models, for each trait.

The GWAS analysis were performed for the different BLUEs datasets: VCU core trials across different environments and DUS trial.

4.2.10 QTLs gene interval exploration

The confidence interval for each QTL was calculated using the LD decay threshold at r^2 of 0.3. Basically, the LD decay values in Mb was added in both directions from the most associated SNP marker of each peak. The confidence interval for each peak was analyzed in Ensembl plant database using the Biomart tool (Bolser et al., 2016), downloading the genes from *Triticum turgidum* cv Svevo RefSeq v1.0 (Maccaferri et al., 2019b). The gene network was studied obtaining the orthologues on Triticum aestivum cv Chinese Spring v1.0 genome, and the gene network was enquired using the knetminer database (Hassani-Pak et al., 2021).

4.3 Results

4.3.1 SNP analysis and LD decay

Genomestudio v2.0 was used to call SNPs by the Polygentrain algorithm, giving good results in terms of clustering and the number of SNPs filtered for suitable quality parameters. However, as described in the material and methods section (chapter 4.2.7), it was decided to further filter the SNPs to the correctly mapped on *Triticum turgidum* cv. Svevo genome before performing imputation. Beagle 5.4 was used for SNPs imputation, and a comparison between the common varieties between durum Panel 1 and 2, Global durum panel (GDP), and Innovar durum wheat DUS dataset was performed, removing SNPs that gave contrasting calls.
Polymorphic information content (PIC) was calculated for the imputed hapmap file (Figure 6), where a major part of markers has a 0.5 allele frequency, meaning they represent informative molecular markers.

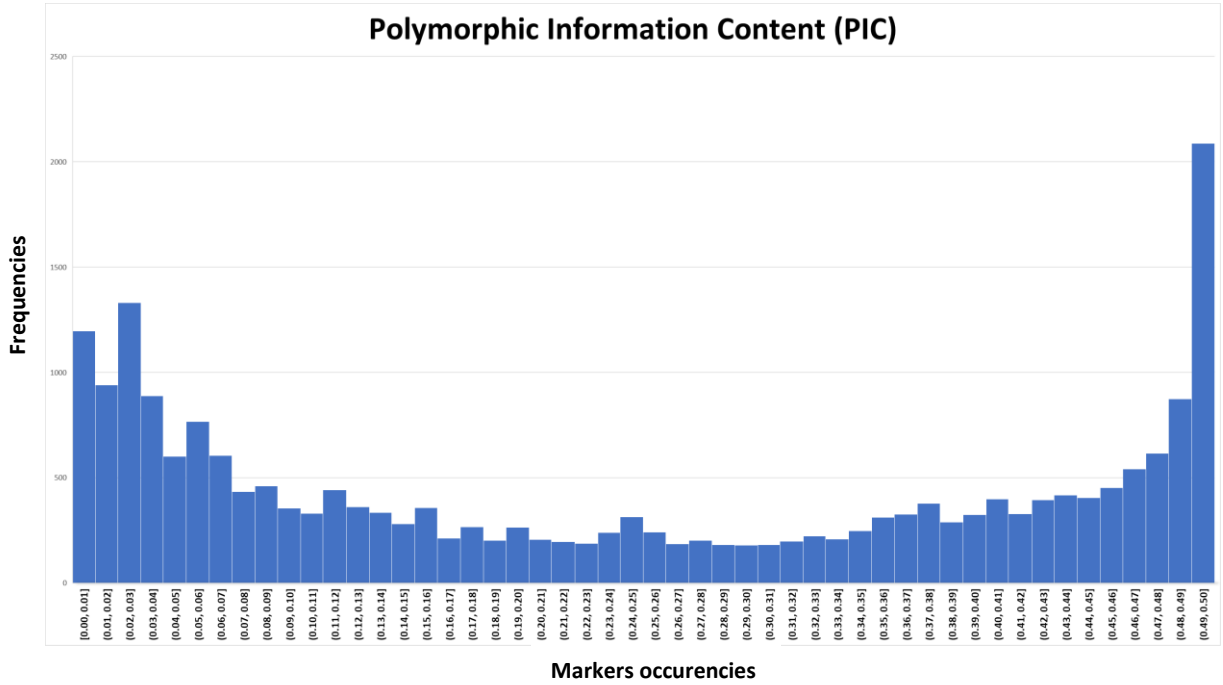


Figure 6: PIC table showing the markers occurrences (y-axis) and the frequencies (x-axis)

LD decay was calculated for the panel at the three r^2 values as reported in Figure 7. For $r^2 = 0.3$ the physical distance was 0.7 Mb, for $r^2 = 0.5$ it was 0.3 Mb and for $r^2 = 0.8$ it was 0.07. Regarding the $r^2 = 0.3$ threshold, the distance value was expected to be greater than 1 Mb as average LD decay reported for modern germplasm collection was reported to be higher (Mazzucotelli et al., 2020). However, the increase in recombination rate can be ascribed to the presence of some landraces related to the modern germplasm in the GDP panel with higher recombination than the modern elite varieties.

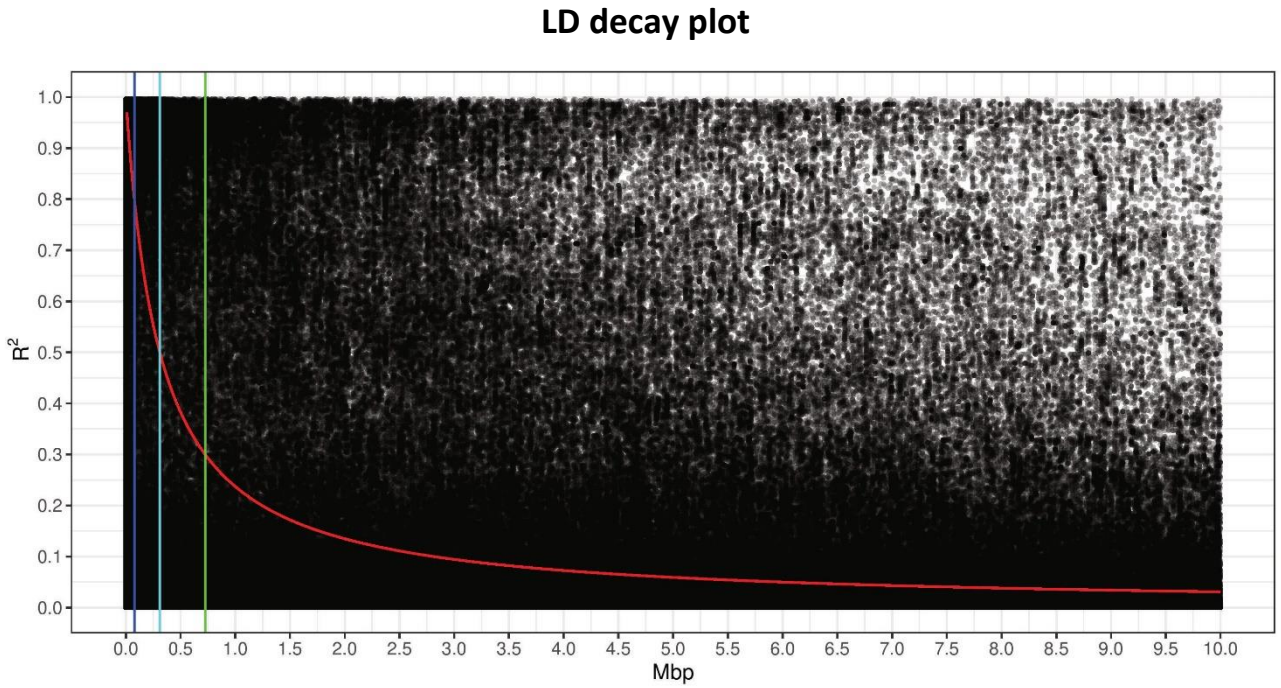
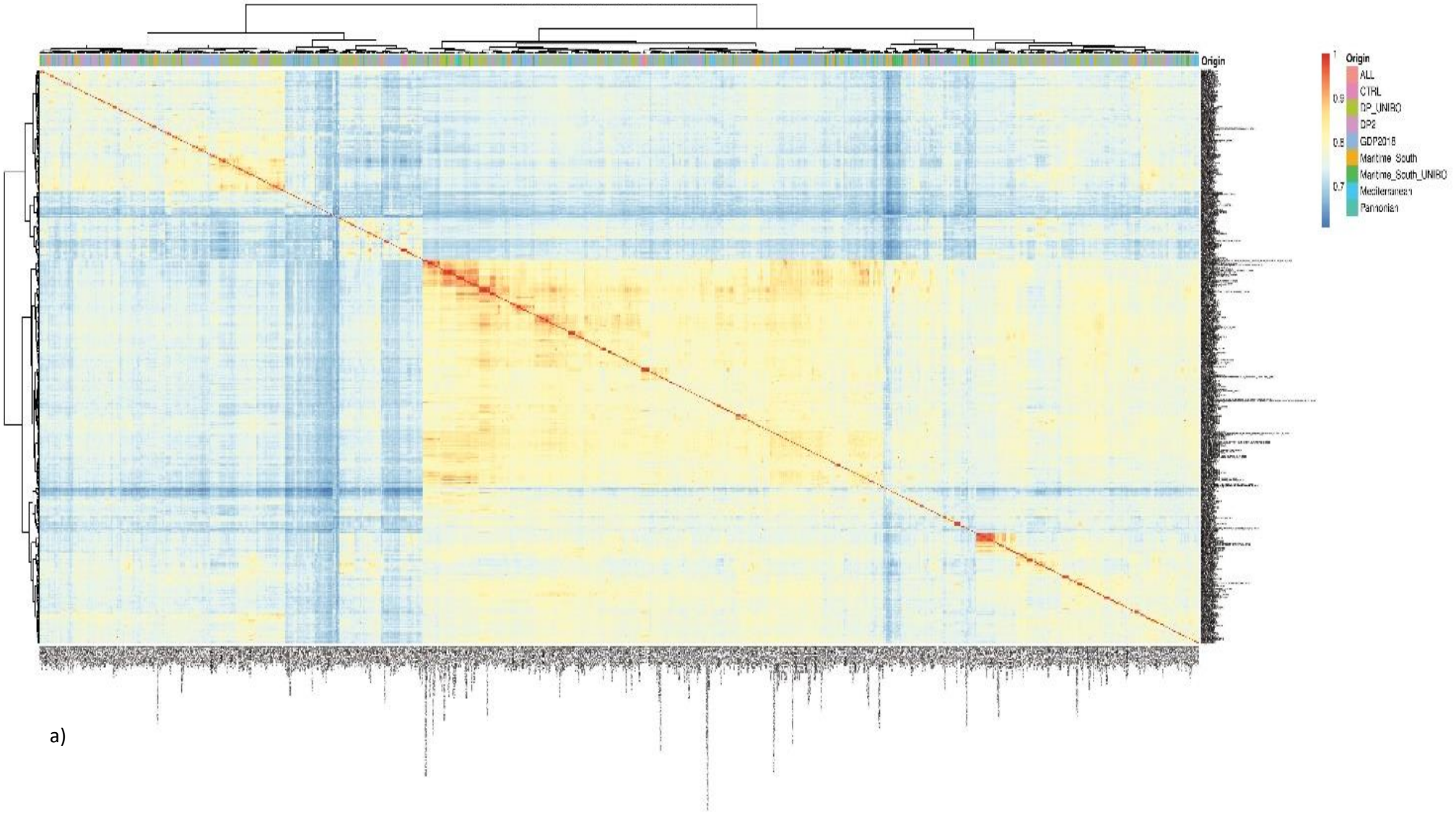


Figure 7: LD decay plot of INN-GDP-DP2-DP2 collection. The vertical lines represent r2 = 0.3 (green), r2 = 0.5 (light blue) and r2 = 0.8 (blue). The r^2 values are reported on the y axis and the Mb units on the x-axis.

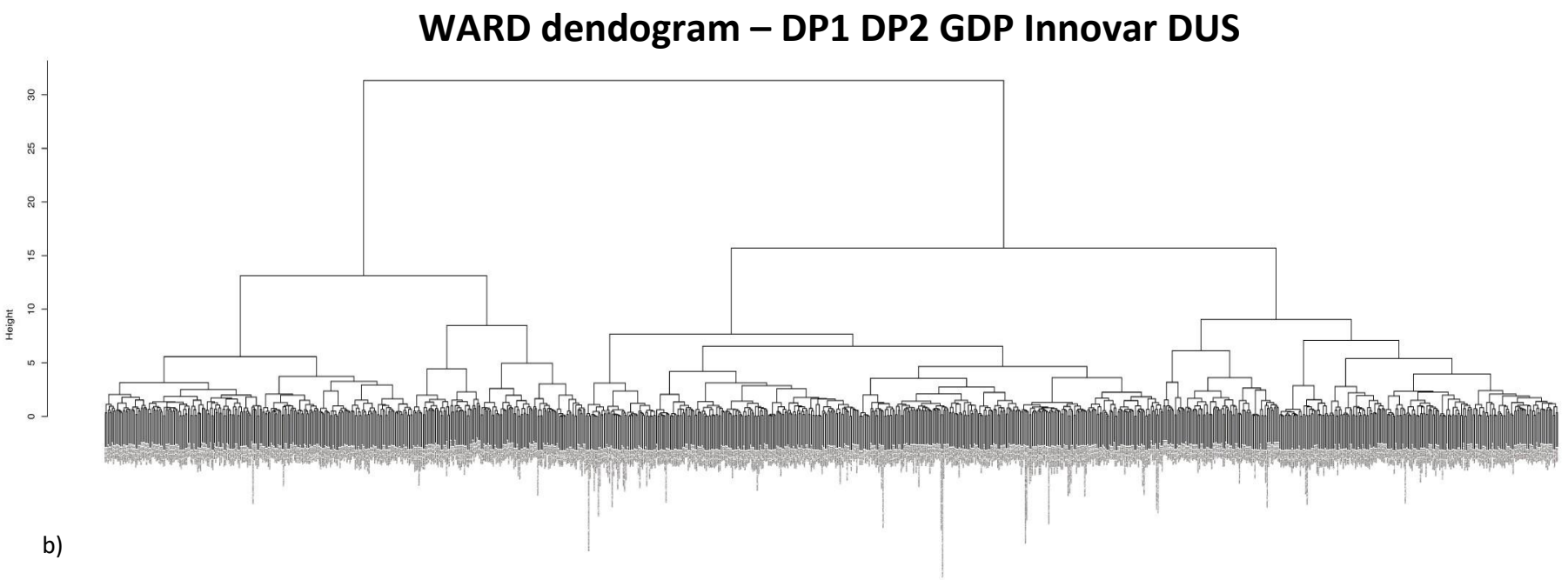
4.3.2 Ancestry analysis

Distance and kinship matrix was computed with TASSEL5, and dendrogram tree was performed with Ward.D2 algorithm, results are plotted in Figure 8 *a* and *b*.

Kinship matrix – DP1 DP2 GDP Innovar DUS panels



a)



b)

Figure 8 a) and b): a) kinship matrix heatmap imputed, different colours are reported as columns annotations based on the origin of the accessions (Innovar ACZ). The kinship matrix was computed using TASSEL5 starting from the distance matrix. b) WARD dendogram as reported in the kinship matrix.

The kinship matrix pointed out the different origins of varieties (4 ACZ INNOVAR, durum panel 1, durum panel 2, plate controls Svevo and Cappelli, and GDP) which can be recognized in specific genetic similarities clusters. As highlighted in the annotation columns on the heatmap, the clusters are composed of related varieties as the ACZ origins are merged among all the genotypes. In fact, varieties were obtained from different breeding programs that used common founders/parents in several cases. Therefore, the genetic distance among the 3 ACZ (Maritime South, Mediterranean, and Pannonian) was not high for a major part of varieties. The highest genetic distances were detected for GDP landraces or wild wheat parental lines included in the collection, which were very different from modern elite durum varieties included in the Innovar durum panel. This was not unexpected as the panel comprises modern accessions; even if adapted to different ACZ they could share common founders and thus be related among them.
The ward clustering showed the genetic relationship among varieties, including common origin/breeding programs reported in Table 5. Close clusters are represented by CYMMIT/ICARDA related breeding lines and Southern Italian varieties, and other clusters contain Northern Italy/central France varieties. In addition, Russian and Turkish landraces principally cluster in separate groups being different from modern cultivars.
Admixture analysis was performed as reported in materials and methods (Chapter 4.2.8), considering populations from K=2 to K=20. To choose the correct number of subpopulations to work with, the different parameters taken into consideration are reported in Figure 9.

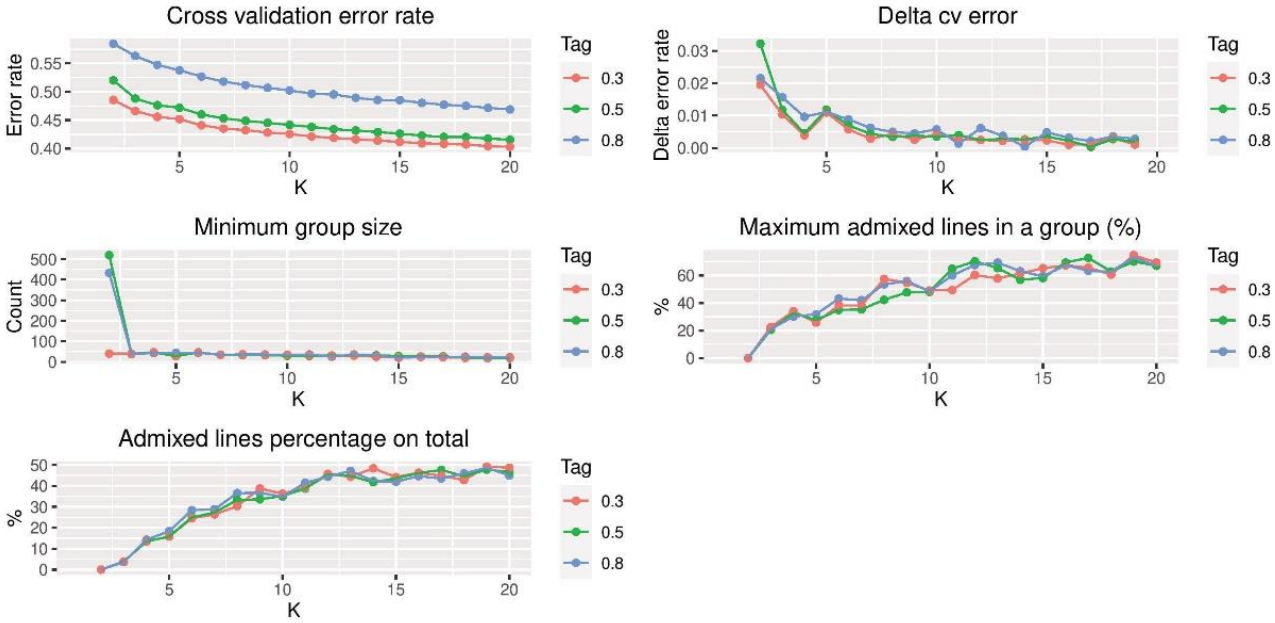


Figure 9: admixture parameters reporting cross validation error rate, delta cv error, minimum group size, maximum admixed lines in a group and admixed lines percentage on total. The values for each one of the three r^2 thresholds (0.3, 0.5 and 0.8) are reported.

The Admixture parameters chosen were k=10 with the r^2 threshold equal to 0.5. These parameters were considered the ones with the lowest error rates and with the more stable admixed lines and groups to be represented (Figure 10).

Admixture plot at k10

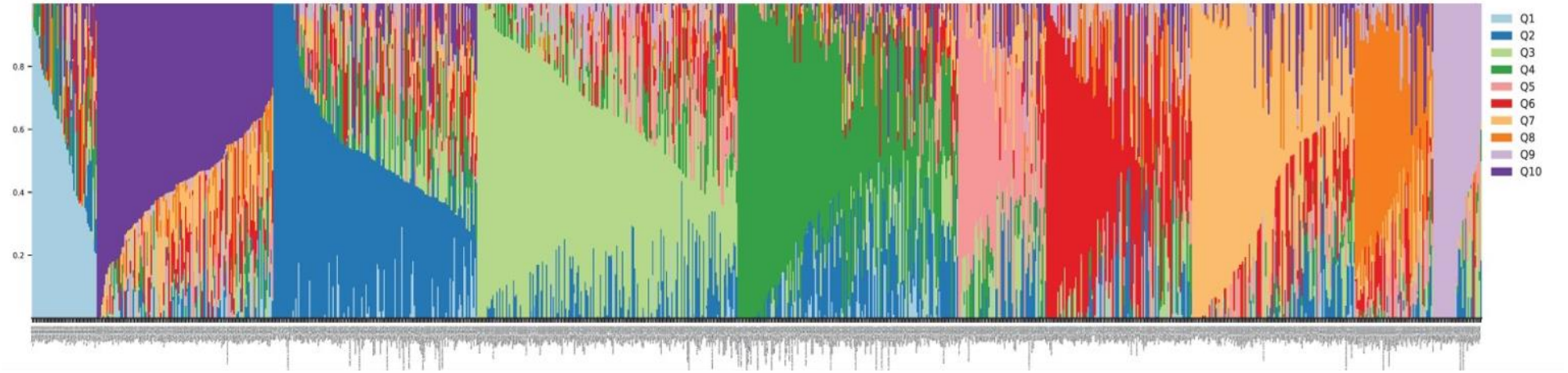


Figure 10: admixture plot at k10 and r^2 0.5. Different clusters are ordered based on Q sub-populations. Varieties are ordered following the hapmap and the number of subpopulations, from 1 to 10.

The herein reported admixture plot reports the different subpopulations ordered from Q=1 to Q=10 detected at k=10. Although, as clearly visible, the percentage of admixed lines is very high, this was not totally unexpected as the germplasm object of study is composed of modern varieties closely related and inbred coming from different breeding programs (Table 5). In particular, the most recent varieties showed a high percentage of admixture reflecting the three-ways complex crosses frequently carried out by breeders.

Table 5: different breeding programs associated to the panel accessions. These programs are clearly determined both in ward and admixture clusters

Breeding programs	**Breeding companies**
CentralFrance	French-Benoist-GAE
CentralFrance_NorthItaly	SYNGENTA-PSB-APSOV-SIS
cimmyt60-earlyICARDATemperate(Cocorit71_Cham1)	CIMMYT-ICARDATemperate60-70
CIMMYT70(YavarosC79_Karim_Duilio_ICARDAtemperate)	CIMMYT
CIMMYT70_related	cimmyt-spanish_breeding
CIMMYT70_Svevo	Italy-Spain
CIMMYT80(Altar84)	CIMMYT
CIMMYT80/90_related	Spanish_Italian_programs
CIMMYT80_related	SYNGENTA-PSB-APSOV-SIS
CIMMYT90_recent	CIMMYT90
DesertDurum	DesertDurum
French_breeding	french_breeding
French_Program(Nefer)	Nickerson_France/APSOV
ICARDA_temperate_recent	ICARDA
ICARDA-Dryland(Haurani_Omrabi_Syrian_Gidara)	ICARDA
LandracesRussian_varieties-Pannonian	Russia_Hungarian
LandracesTurkishMediterranean	LandracesTurkishMediterranean
NorthDakota-French	NorthDakota_Program
NorthernItaly	SYNGENTA-PSB-APSOV-SIS
Southern_Italy_breeding70(Valnova_Mexicali_Grazia)	Italian_Breeding
SouthernItaly group (Creso founders)	Italian_Breeding
SouthernItaly group (Valnova-Capeiti-Cappelli founders)	Italian_Breeding
SouthernItaly group (Valnova-Capeiti-Cappelli founders)	Italian_Breeding
Turkey-CentralAsianLandraces-CentralFrance_Austrian_Canadian	Admixed complex cross and landraces
various_program_highAdmixture	Admixed complex_cross

The admixture cluster results were compared to the ward cluster reported by the kinship matrix, using the same colours as in the admixture plot (Figure 11). The important note to point out is that the ward dendrogram allows recognizing more specific breeding groups and better separates the different relationships between varieties. Admixture software can detect different clusters as well, but, probably, as the percentage of admixed lines in modern cultivars is very high, it does not mirror perfectly what is shown by the ward dendrogram.

Kinship matrix grouped with Admixture subpopulations

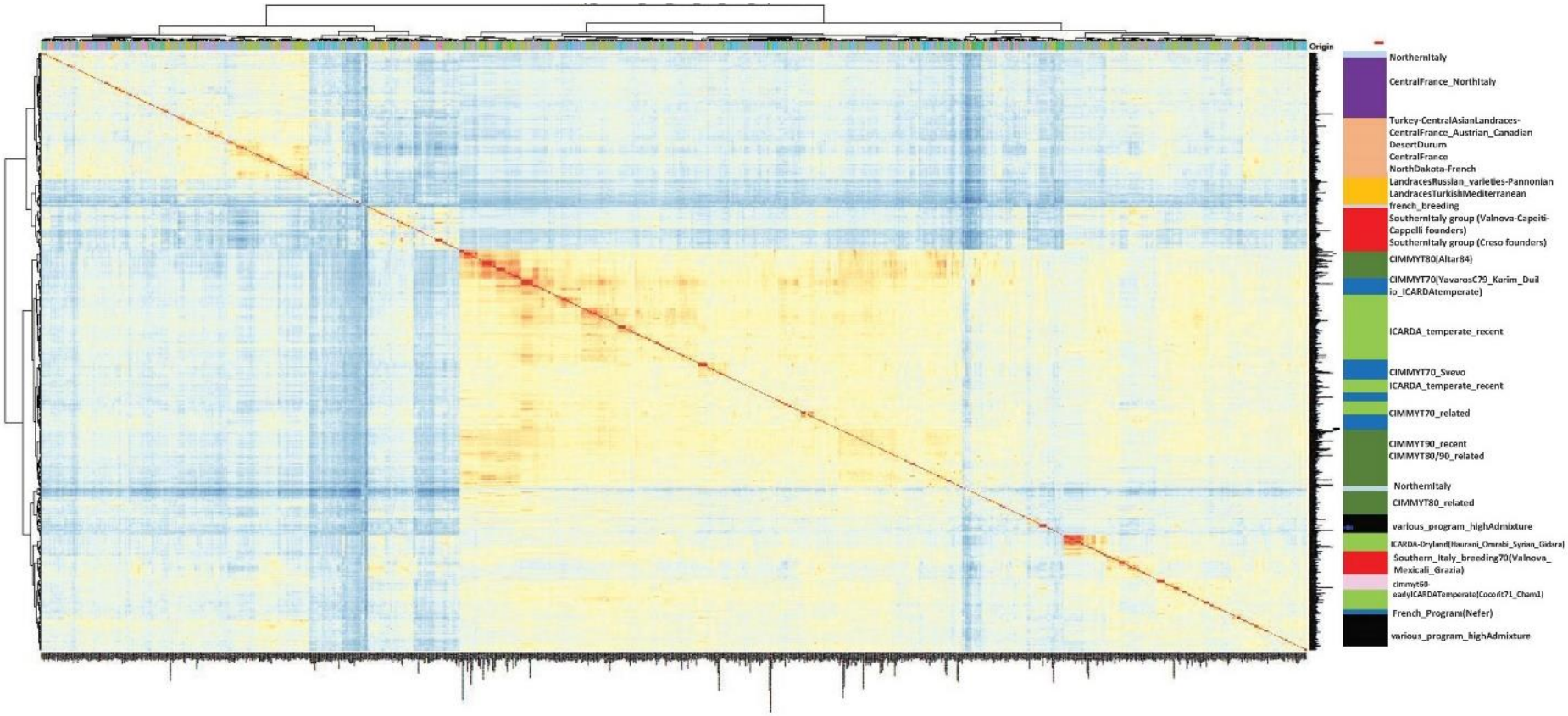


Figure 11: kinship matrix, the column on the right reports the correspondence between the different groups detected by admixture. Each colour represents a breeding program (Table 4) with the same colour palette reported in the admixture plot.

The distance matrix computed with TASSEL5 was used to calculate the Neighbour joining tree clustering among all the varieties in the panel (Figure 12). The tree shows how CYMMIT, ICARDA, and Southern Italian varieties tend to cluster together. On the other hand, the northern Italian and French varieties represent close clusters, and Russian and Turkish landraces are more separated in different clades. However, a certain degree of admixture between different clades is still visible given the modern origin of the major part of varieties.

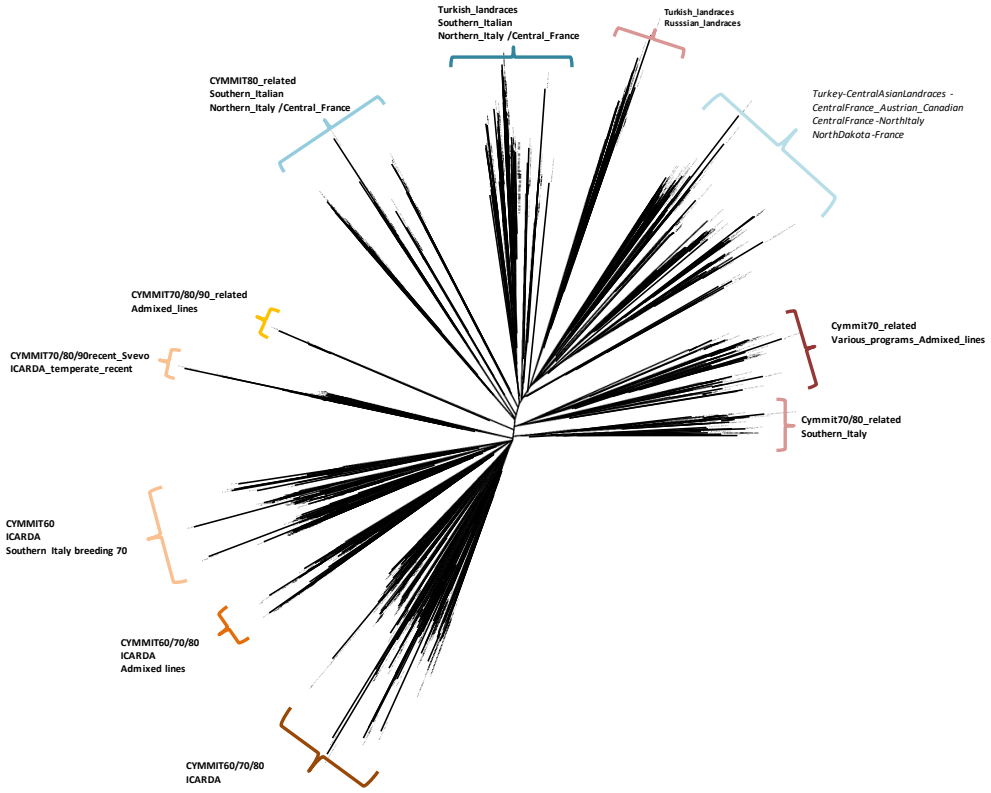


Figure 12: Neighbor Joining Tree of the complete panel. The clades are highlighted based on the origin of the corresponding breeding program

4.3.3 Phenotypic analysis on VCU field trials

The VCU phenotypic traits were statistically analysed with R and Rstudio softwares and different phenotypes were evaluated for distributions and significances between the treatments. Considering all the performed VCU trials, correlation analysis between different phenotypic traits was performed (Figure 13).

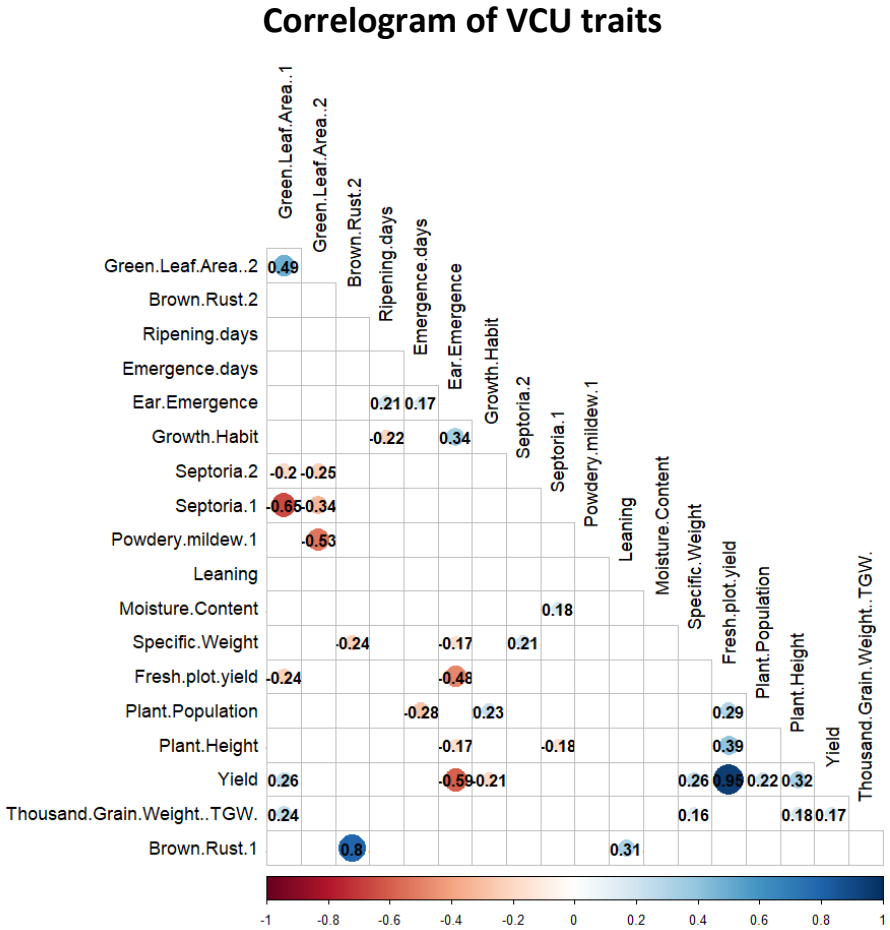


Figure 13: correlation plots of the VCU phenotypic data. Only significant correlation values (p-value = 0.05) are reported. Diseases were scored on different dates, distinguished by .1 and .2 at the end of the phenotypic trait name.

Based on the correlogram, only significant correlation values were reported based on p-value of 0.05 as statistically significant threshold. As expected, a positive correlation was detected between yield and traits with a strong phenological connection, such as fresh plot yield, thousand kernel weight (TGW), plant height and specific weight. Negative correlation was detected between yield related traits and different diseases, as expected. Furthermore, yield was negative correlated with ear emergence (heading date). This probably shows that the later the heading date is, the more affected the plant is to different foliar pests and diseases, thus reducing the final yield production. In general, VCU protocol includes the evaluation of different diseases across the different environments. However, the disease pressure was not high enough to be statistically evaluated and to compute GxE interaction across different environments. Therefore, data analysis and GWAS were principally focused on yield related traits on VCU core and drought trials, as the disease percentage did not significantly affect the final yield production (especially on VCU core trial). These results are

also connected with the different fungicide treatments that were performed on VCU core trial, namely full and minimal treatment levels (Figure 14 a). Due to the low disease pressure, the total yield quantity (kg/ha) was not strongly significantly different between full and minimal treatment in VCU durum core trials.
The same results were consistent with the yield production divided for the different environments. However, lower values were detected for UNITUS (Italy) environment as the field trial was subjected to strong cold stress during the growing season. On the other hand, the average yield production was not strongly different across the other environments (UNIDEB = Hungary, HORTA, UPM = Spain) (Figure 14 b).
Similar results were obtained also for VCU drought trials, where fungicide treatments were homogeneous in the field trial, but two different drought regimes were applied: irrigated and not irrigated. As expected, the irrigated treatments had a higher total yield compared to the not irrigated treatments (Figure 14 c).
The comparison between the VCU drought trial across different environments show similar results to the VCU core trials, namely an average similar yield with the only exception of UNITUS as it was subjected to cold stress during growing season (Figure 14 c and d).

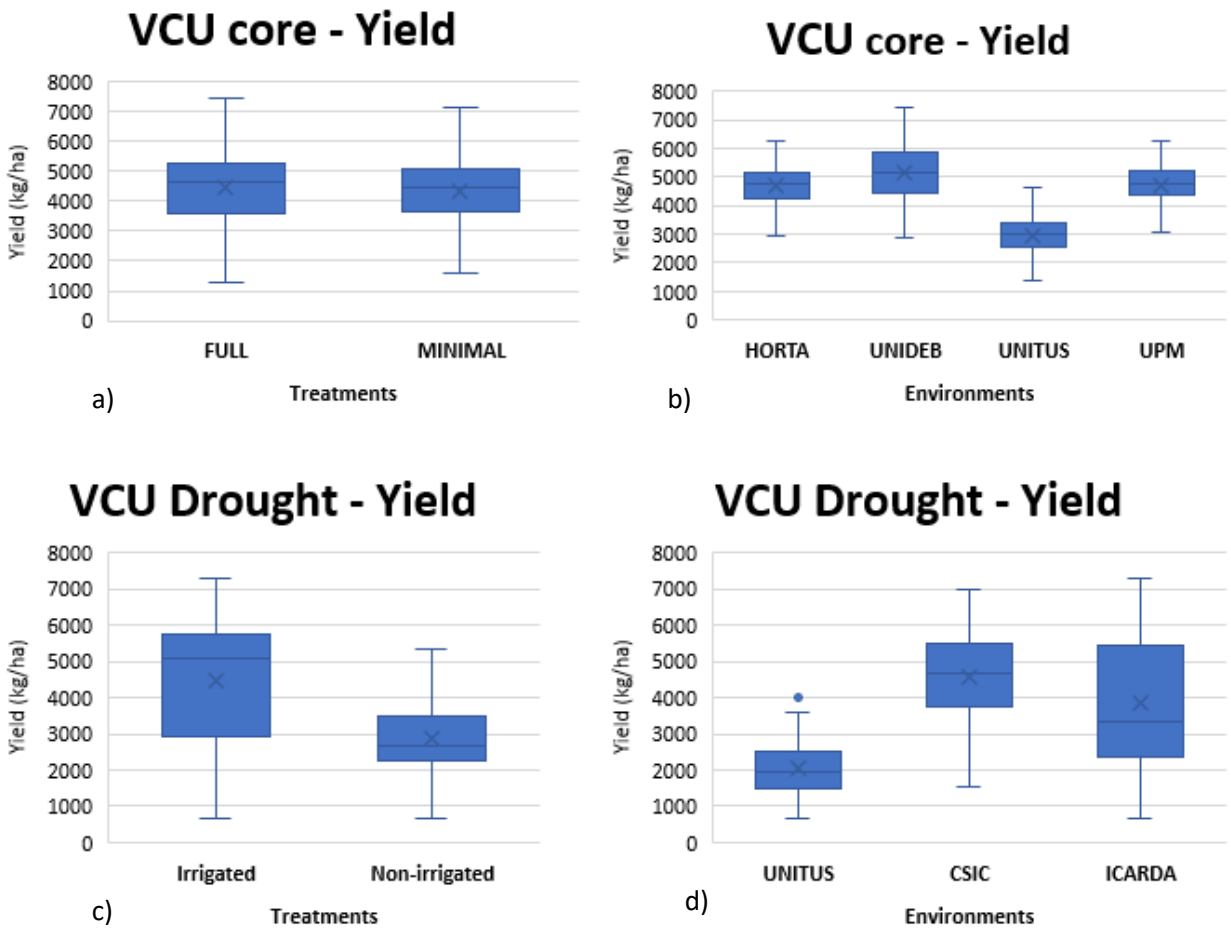


Figure 14: a) and b) yield boxplots on VCU core trials across different treatments and environments. The final yield was plotted between the two fungicide treatments, full and minimal. The yield was calculated in kg/ha. c) and d) report yield boxplots of VCU drought trial between irrigated and not-irrigated treatments and across different environments.

The following phenotypic analysis was performed on VCU core trials; however, it was carried out also in VCU drought trials but data are not shown due to the similar values detected between core and drought trials.
VCU core trial data regarding heading date (measured as number of days from the sowing date), fresh plot yield (kg/ha), plant height (cm), thousand grain weight (TGW) (g) and yield corrected for the 15% of moisture content (kg/ha) were analysed obtaining descriptive statistic and heritability values which ranges from 0.73 to 0.95 (Table 6).

Table 6: descriptive statistics of the VCU main traits studied in VCU core trial.

	Heading date	**Fresh_plot_yield**	**Plant_height**	**TGW**	**Yield_15**
min	145.96	2.62	53.31	28.38	2901.82
max	166.42	5.57	85.31	49.56	5847.06
range	20.46	2.94	32.00	21.18	2945.25
median	154.08	4.36	74.20	39.92	4696.52
mean	154.44	4.26	73.72	40.24	4628.44
Standard error	0.35	0.06	0.52	0.36	49.55
variance	16.80	0.29	28.39	17.15	333889.62
Standard deviation	4.10	0.54	5.33	4.14	577.83
Coefficient of variation (CV)	0.03	0.13	0.07	0.10	0.12
h^2	0.95	0.79	0.91	0.86	0.73

The phenotypic traits reported in Table 6 show the consistency of the measured data on the VCU core trials, based on the low CV and on the high heritability values.
This stress the genetic component of the yield related traits considered in respect to the environmental influence.
The phenotypic distribution of the yield related traits, after BLUES extraction, didn't correspond to a normal distribution based on the high sensitivity of the shapiro test (data not shown). However, it was anyway considered similar to a normal curve (Figure 15) considering the average curve shape.

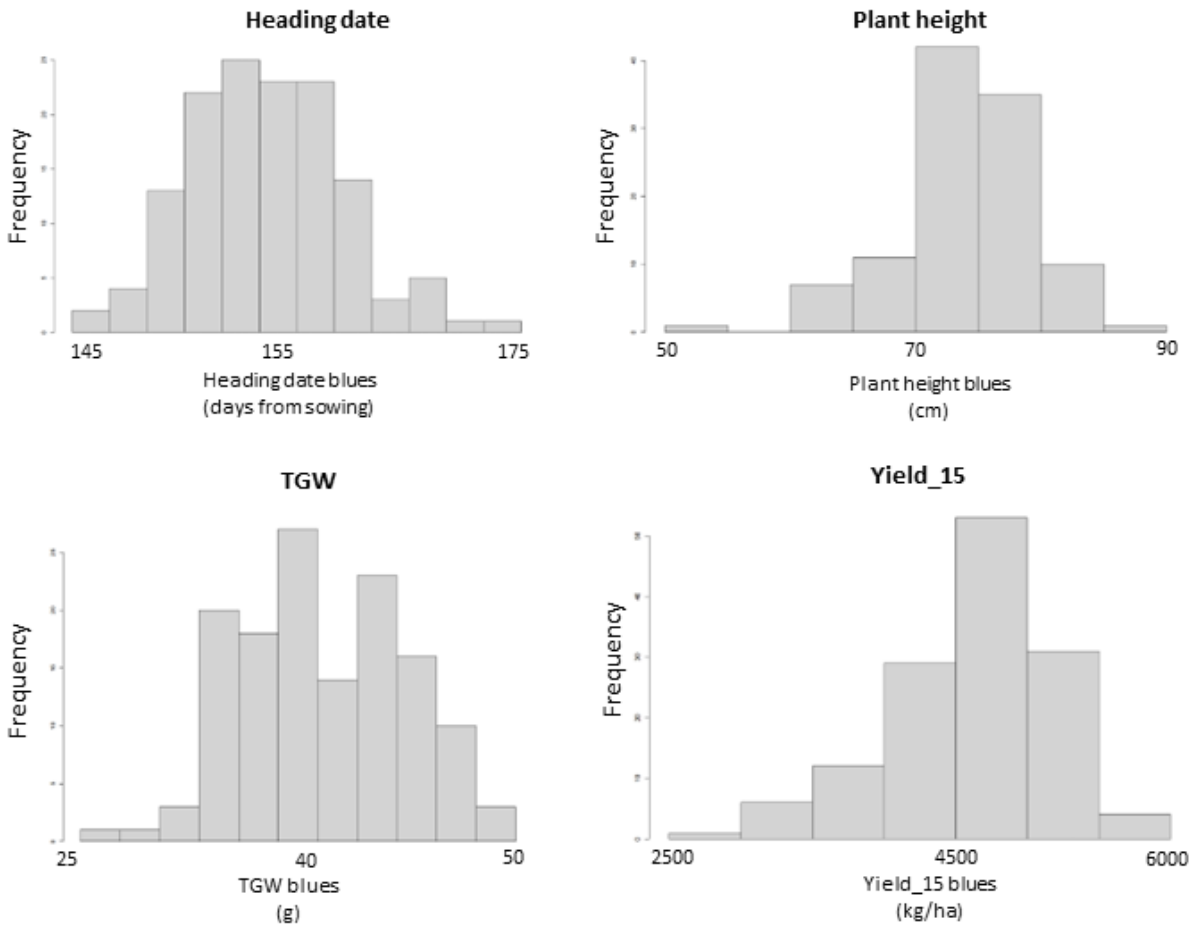


Figure 15: phenotypic distribution of the main phenotypic traits: heading date, plant height, TGW and total yield corrected for 15% of moisture content. The distributions seem to reflect a normal curve

The ANOVA analysis showed the interaction of genotypes between treatments and environments for VCU core trials, identifying significance of the E component (Environment) and the interaction GxE between genotypes x treatments, but also between environments x treatments and replicates x treatments (Table 7).

*Table 7: ANOVA tables referring to the VCU Core trials.Columns reports the trait, the considered variable, the sum of squares, the degrees of freedom, the F values and the p-values. Significance values reported are: < 0.1= ., <0.05 = ** and < 0.01 = ***.*

Trait	Variables	Sum Sq	Df	F value	p-value	
Plant height	**Genotype**	19612.4	106	9.7539	2.20E-16	***
Plant height	**Replicate**	1089.5	2	28.7167	1.10E-12	***
Plant height	**Environment**	30801.3	2	811.8864	2.20E-16	***
Plant height	**Treatment**	323.9	1	17.0751	4.06E-05	***
Plant height	**Genotype x Environment**	2798	24	6.1461	2.20E-16	***
Plant height	**Environment x Treatment**	573.6	2	15.119	3.81E-07	***
Plant height	**Replicate x Treatment**	41.8	2	1.1029	0.3325	
Plant height	**Residuals**	12481.6	658			
Heading date	**Genotype**	15010	135	44.4882	2.20E-16	***
Heading date	**Replicate**	13	2	2.5347	0.07996	.
Heading date	**Environment**	323002	3	43079.34	2.20E-16	***
Heading date	**Treatment**	29	1	11.7282	0.000649	***
Heading date	**Genotype x Environment**	2714	44	24.6777	2.20E-16	***
Heading date	**Environment x Treatment**	5	2	0.9296	0.395151	
Heading date	**Replicate x Treatment**	0	2	0.0992	0.90558	
Heading date	**Residuals**	1879	752			
TGW	**Genotype**	16691.6	135	13.2397	2.20E-16	***
TGW	**Replicate**	25	2	1.341	0.262108	
TGW	**Environment**	29526.4	3	1053.904	2.20E-16	***
TGW	**Treatment**	129.7	1	13.8927	0.000206	***
TGW	**Genotype x Environment**	1796.5	44	4.3721	2.20E-16	***
TGW	**Environment x Treatment**	99.3	3	3.5431	0.014277	*
TGW	**Replicate x Treatment**	26.1	2	1.3972	0.247826	
TGW	**Residuals**	8376.9	897			
Yield.15	**Genotype**	3.23E+08	135	4.1091	2.20E-16	***
Yield.15	**Replicate**	24698273	2	21.1982	1.01E-09	***
Yield.15	**Environment**	3.08E+08	3	176.1856	2.20E-16	***
Yield.15	**Treatment**	4574145	1	7.8519	0.005185	**
Yield.15	**Genotype x Environment**	85175293	44	3.3229	9.93E-12	***
Yield.15	**Environment x Treatment**	8265638	3	4.7295	0.00279	**
Yield.15	**Replicate x Treatment**	983390	2	0.844	0.430311	
Yield.15	**Residuals**	5.28E+08	907			

In conclusion, the yield related traits show a normal distribution and high heritability values showing consistency in data acquisition across different VCU core trials. The ANOVA analysis show the significance of environmental variable in the phenotypic evaluation. The models were included with the analysed variables in ANOVA tables (Table 7) and BLUES were extracted in order to be used as phenotypic corrected data to perform GWAS analysis.

4.3.4 Phenotypic analysis on DUS field trial

The DUS trial was carried out in in collaboration with CREA-DC (Italy), monitoring and detecting different morphological traits based on the EU DUS protocol. The correlation plot showed positive correlation between flag leaf traits, heading date with plant height, ears and grain traits, ears and awn colors and negative correlation of these two traits with glaucosity features (Figure 16).

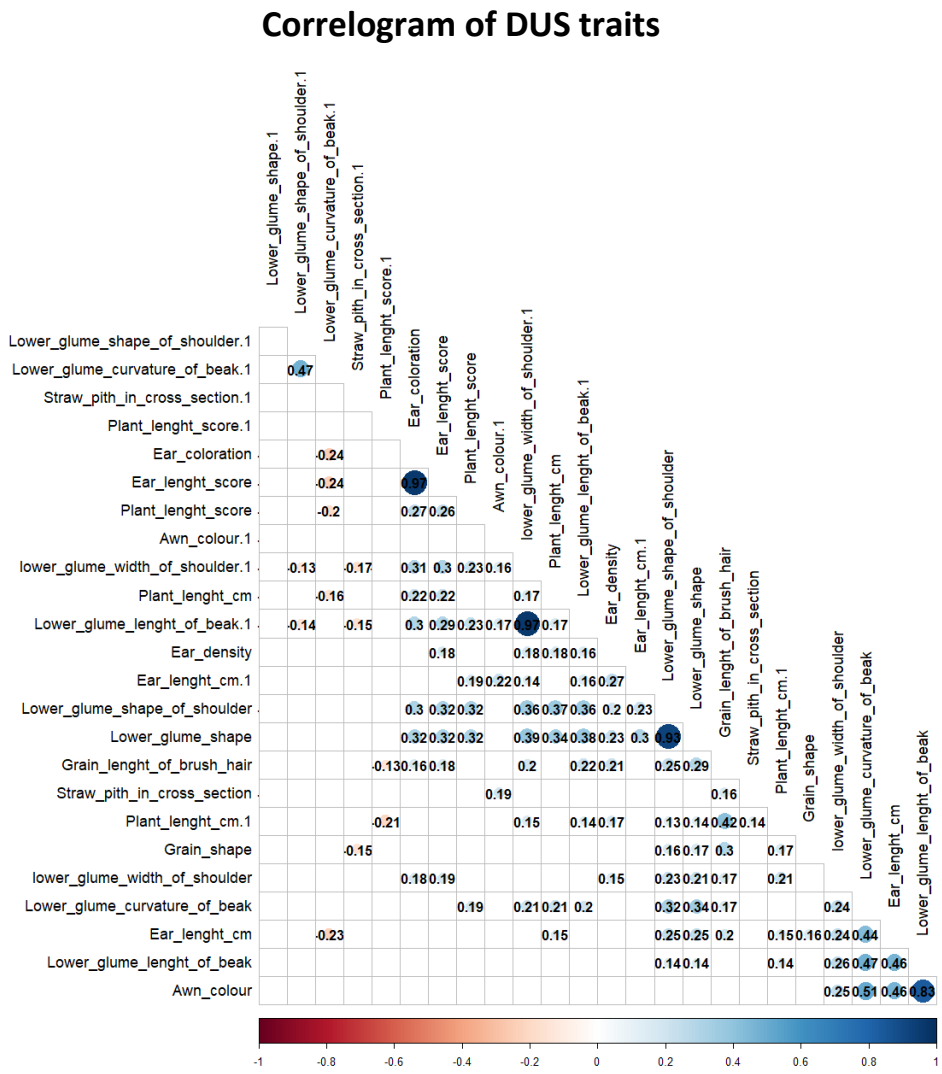


Figure 16: correlation plots of the DUS phenotypic traits. Only significant correlation values were included, (p-value threshold=0.05).

The DUS field trials took in considerations different morphological features. However, few main characters were considered and will be treated in detail, such as: plant growth habit, heading date,

ear and culm glaucosity, ear and awn colours. The descriptive statistics of the main DUS traits show, as reported for the VCU, the consistency of the data based also on the high heritability of the trait, that ranges between 0.63 and 0.99. These data report the high genetic component of these traits inside the DUS in comparison to the environmental variables (Table 8).

Table 8: descriptive statistics and heritability values of the main DUS traits taken in consideration.

	Plant growth habit	**Heading date**	**Culm glaucosity of neck**	**Ear glaucosity**	**Ear density**	**Awn colour**	**Ear colour**
min	1.00	182.47	6.73	2.90	3.99	0.98	1.00
max	4.03	197.6	9.22	9.08	7.01	4.02	3.00
range	3.03	174.14	2.48	6.18	3.02	3.04	2.00
median	2.50	189.93	8.05	7.01	6.00	2.00	2.00
mean	2.51	190.38	8.11	6.86	5.68	1.96	1.73
Standard error	0.04	0.22	0.04	0.08	0.05	0.05	0.03
variance	0.36	11.05	0.32	1.40	0.57	0.71	0.28
Standard deviation	0.60	3.32	0.56	1.18	0.76	0.84	0.53
CV	0.24	0.11	0.07	0.17	0.13	0.43	0.30
h^2	0.66	0.96	0.63	0.85	0.98	0.99	0.99

The phenotypic distribution of the traits is not normal based on shapiro test (data not shown) due to the high-test sensitivity. However, based on the curve shape, it may also be considered as a kind of normal distribution for a major part of the morphological traits monitored in DUS trial (Figure 17). The only exceptions are represented by the ear coloration, whose distribution can be more similar to a bimodal curve, hypothesizing a strong dominant QTL responsible for the regulation of these traits.

As the DUS trial was composed only of one environment, phenotypic data in the DUS trial were corrected for the replicates (environment) but also for spatial distribution of each variety in the field expressed as position coordinates with rows and columns (Table S15).

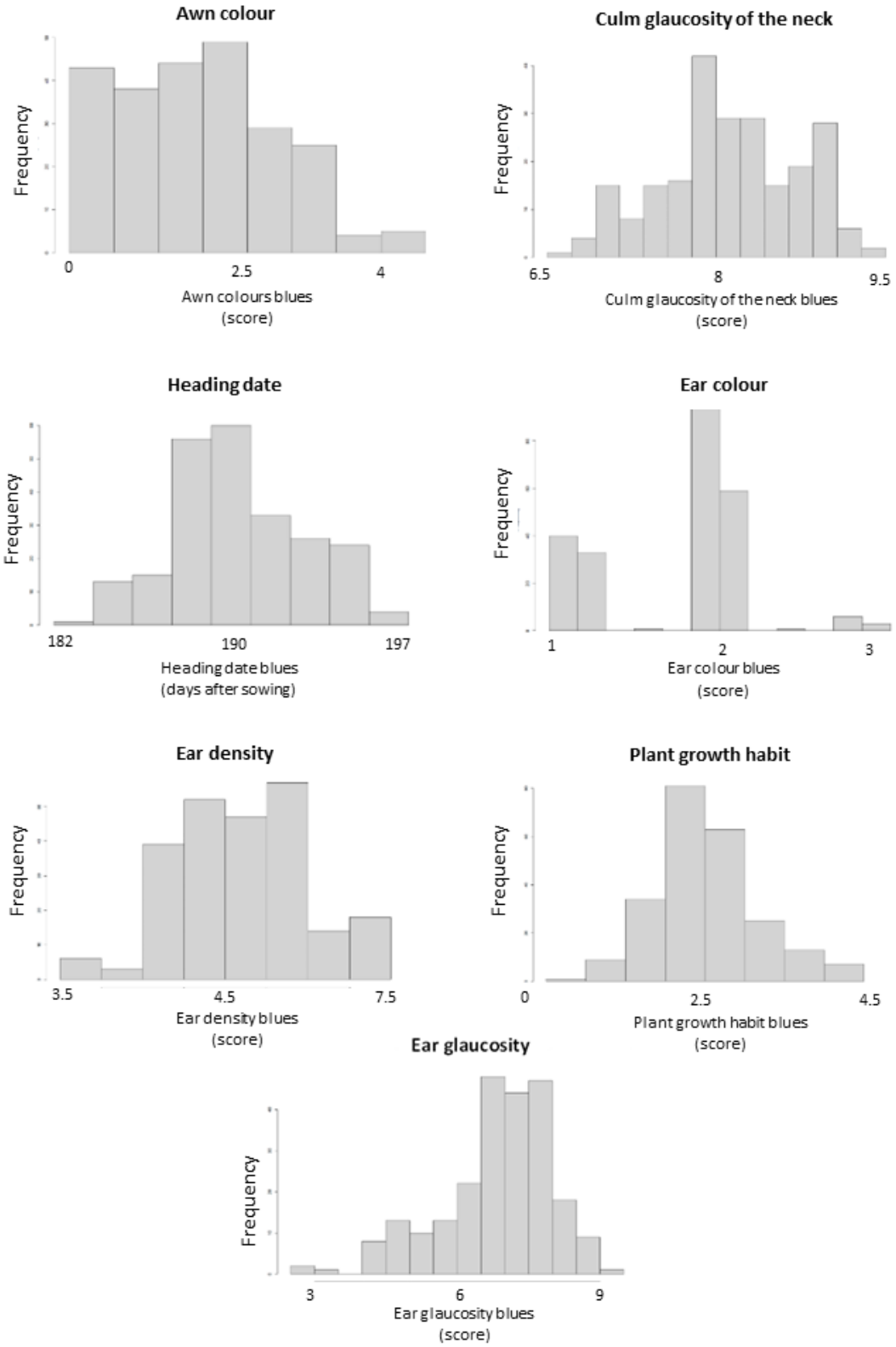


Figure 17: phenotypic distributions of the DUS main traits: heading date, plant growth habit, ear density, culm glaucosity of the neck, ear glaucosity, awn and ear colour.

4.3.5 VCU GWAS results

Once that the BLUES data for the VCU core and drought trial were obtained (using the merged dataset for all the environments), GWAS was performed using GAPIT3 software. The analysis was performed using different models (correcting for population structure with kinship matrix), namely: GLM, MLM, MLMM, FarmCPU and BLINK. Starting from the original hapmap (already imputed) filtered for the correctly mapped SNPs, which included also accession from GDP, DP1 and DP2, a filtering step was performed to retain only varieties included in the VCU field trials, reaching a final variety number of 166 varieties. The VCU hapmap was filtered for MAF (5%) and missing values greater than 50% for SNPs and genotypes. The VCU hapmap was used to compute LD decay and kinship matrix with TASSEL5, population structure analysis with ADMIXTURE. As regards the VCU genotypic data, the computed LD decay at r^2 threshold of 0.3 was approximately 1.34 Mb (Figure 18).

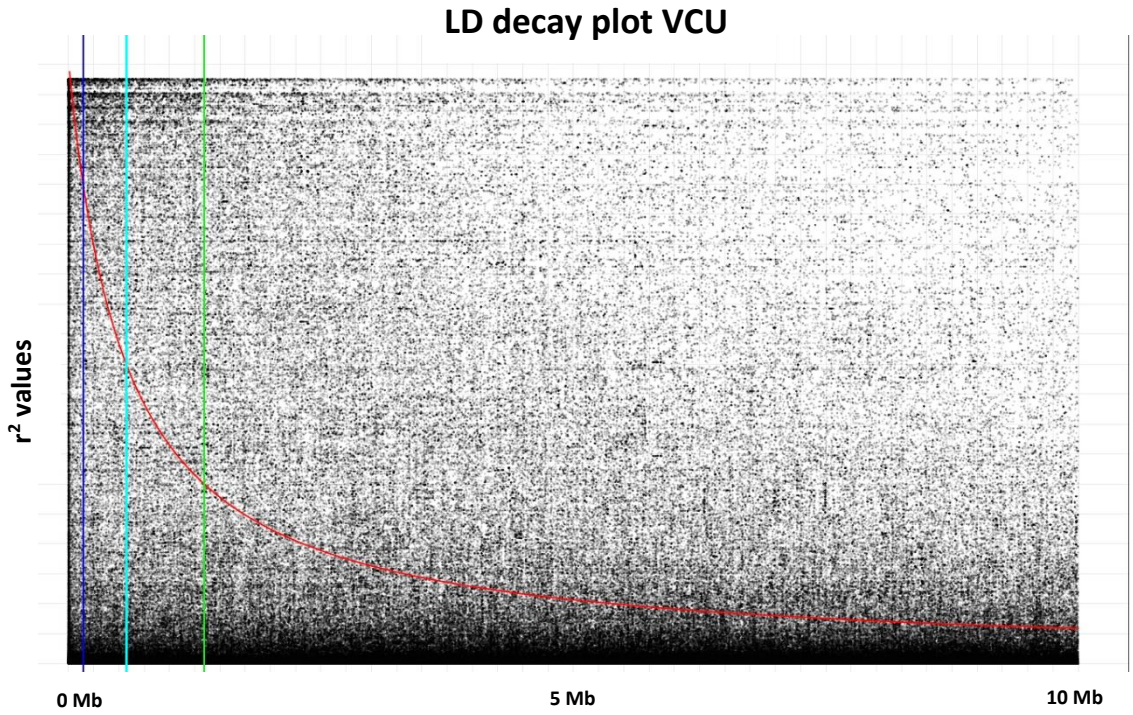


Figure 18: LD decay plot for the VCU genotypic dataset. Three r^2 thresholds were plotted: 0.8 (blue line), 0.5 (light blue line) and 0.3 (green line).

Furthermore, based on the genetic similarities detected in kinship matrix, the average genetic similarity was calculated for each variety across all the VCU panel. This analysis was performed to detect the most genetically distant varieties in comparison to the others included in the VCU trial. As reported in Figure 19, the most similar varieties based on genetic distance are Amilcar, Avispa, Farah, Iride (Avispa haplotype) and Dorondon which come from the CYMMIT-ICARDA breeding program. On the other hand, the most distance varieties are Relief, Lupidur, RGT Voilur, and Monastir which come from European breeding programs (Northern Italy and France) (Figure 19). Taking in consideration the whole population, this result was expected as a major part of the durum varieties used in VCU trial come from ICARDA-CYMMIT or southern Europe breeding programs which lead to higher genetic similarities between these varieties. This result was also reflected by admixture and kinship analysis where there was a high percentage of admixed varieties due to the common origin of the major part of varieties and of most breeding programs.

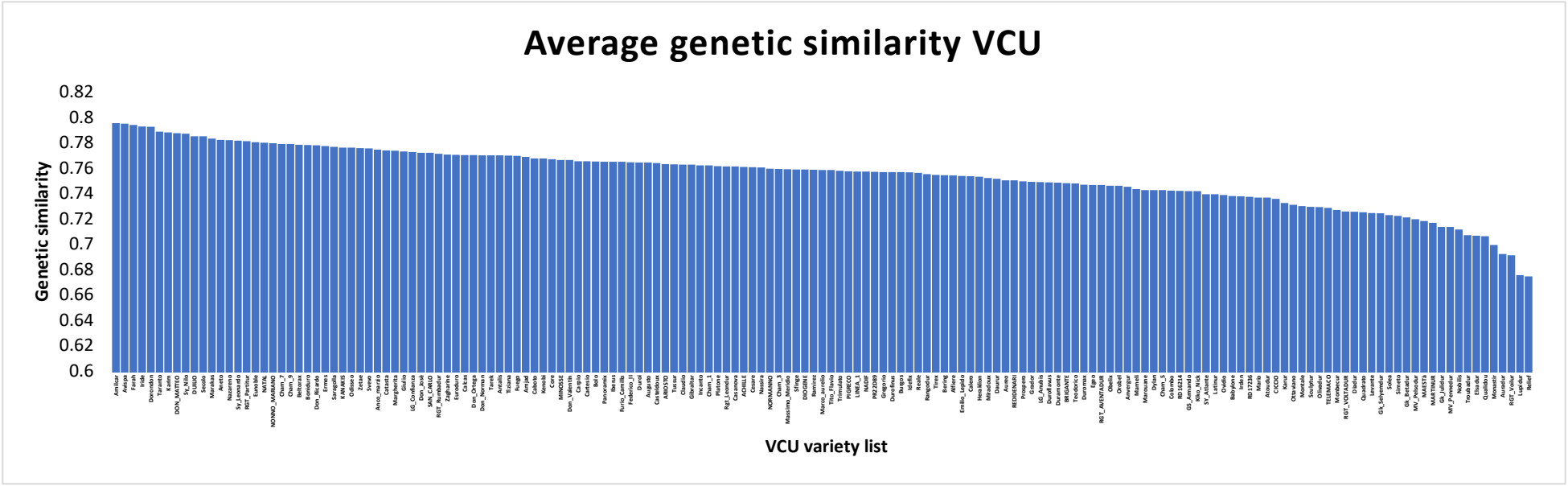


Figure 19: average genetic distance matrix based on kinship of VCU genotypic dataset. The average genetic distance was calculated for each variety and plotted in decrescent order. The x-axis reports the genotypes included in the VCU dataset, the y-axis reports the average genetic distances calculated with identity by state method.

Before computing the different GWAS models, adjusted Bonferroni threshold was calculated performing 1000 permutations with FarmCPU model, which gave similar results than calculating the threshold based on the total number of markers pruned at r^2 threshold equals to 0.8. The calculation was performed dividing the significant p-value threshold (0.05) for the pruned number of SNP markers. In average, the Bonferroni threshold ranged from 4 to 5 and it was used as a cutoff to detect the significant associated markers for each trait. This means that, the considered peaks above the threshold had an increased probability of 10^4-10^5 to associated with phenotypic variance.
The GWAS analysis was divided in two steps. The first pipeline included all the already mentioned model included the kinship matrix in the analysis, with the only exception of the GLM which is the naïve model. The second pipeline uses both the kinship matrix and also the population structure Q file, which represents the value of membership for each variety in every population (from k = 2 to k = 20). In order to calculate the correct number of k (subpopulations), to include in the GWAS, different quality parameters were calculated to choose the number of k with the lowest error rates (Figure 20). As regards the VCU panel, the number of subpopulations that were used for the GWAS analysis was k = 6 at r^2 = 0.5, as, in particular, it showed the lowest cross validation error rate.

Figure 20: admixture quality parameters calculated for the VCU genotyping dataset from k=2 to k=20. The parameters calculated were the cross-validation error rate, the delta cv error, the minimum group size, the maximum admixed lines in a group and the admixed lines percentage on total.

As already mentioned before, the VCU analysis both for core and drought trials was performed on all traits (merging all the environments), with particular interest to yield related traits, which are going to be herein shown, as they are key characters for the VCU protocols (Figure 21). Herein the results on VCU core data across all environments are going to be reported using the pipeline with population structure corrected for the kinship matrix. In fact, correction for kinship and k=6 (population structures), showed the same results as correcting just for kinship, with less QTLs and the major QTLs which were the same in both analysis (background information, data not shown).

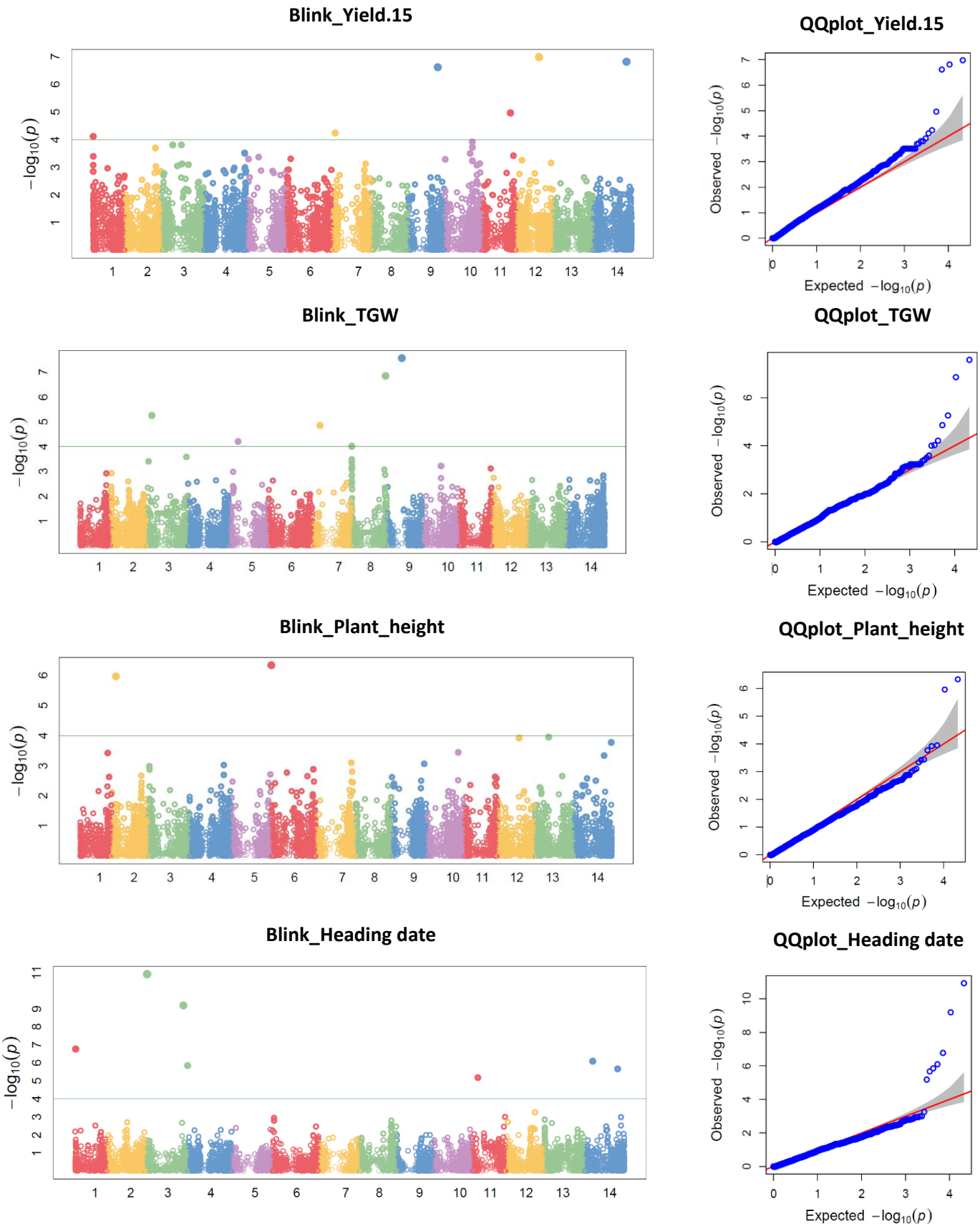


Figure 21: Manhattan plots and QQplots of the GWAS results on yield related traits. The Blink module was shown for all the traits, the threshold was calculated based on the Bonferroni adjusted value.

The Manhattan and QQ plots reported in Figure 21, represent only the BLINK model, even if the the GWAS was performed using GLM, MLM, MLMM, and FarmCPU (Figure S32, Figure S33, Figure S34 and Figure S35). The confidence intervals for each QTL were calculated including the LD decay value (1.34 Mb) at both sides from the most associated SNP for each QTL.
Table 9 reports the associated markers for each trait detected with the BLINK model, mapped on the Svevo RefSeq v1.0 (Maccaferri et al., 2019b) reference genome and the R^2 SNP effect in relation to the phenotype and the effect of the minor allele on the major allele for each SNP.

Table 9: most associated SNP based on GWAS results for each trait are reported. The main peaks considered from major QTLs are underlines. For each trait the confidence interval (C. I.), the R^2 linear effect and the effect of the minor allele on the major allele based on the GLM model are reported.

Trait	SNP	Chr	Position	-logP	C.I (+/-LD decay)	R2_GLM	Effect_GLM
Yield15	IWB49077	12	431988767	6.98	430648767-433328767	0.11	-2.68
Yield15	IWB33975	14	624077880	6.81	622737880-625417880	0.11	2.38
Yield15	IWB9678	9	546814613	6.61	545474613-548154613	0.15	2.03
Yield15	IWA4842	11	519603043	4.96	518263043-520943043	0.01	1.56
Yield15	IWB32978	7	61253977	4.23	59913977-62593977	0.01	-1.96
Yield15	IWB27821	1	3990174	4.11	2650174-5330174	0.06	-1.82
Heading_date	IWB70098	3	37397799	10.92	36057799-38737799	0.28	4.47
Heading_date	IWB68084	3	693970107	9.19	692630107-695310107	0.20	-10.09
Heading_date	IWB3087	1	8966911	6.77	7626911-10306911	0.04	2.00
Heading_date	IWB71644	3	771308796	5.85	769968796-772648796	0.18	3.58
Heading_date	IWB8179	11	103857267	5.18	102517267-105197267	0.05	-2.07
Heading_date	IWB73957	14	145663302	6.09	144323302-147003302	0.23	4.10
Heading_date	IWB22270	14	597479372	5.67	596139372-598819372	0.03	-2.05
TGW	IWB73436	9	279785716	7.57	278445716-281125716	0.12	1.37
TGW	IWB73620	8	646423185	6.85	645083185-647763185	0.21	-1.82
TGW	IWB48585	3	94956501	5.25	93616501-96296501	0.15	-2.12
TGW	IWB47937	7	138502243	4.85	137162243-139842243	0.10	2.14
TGW	IWB67595	5	167185921	4.21	165845921-168525921	0.10	1.29
TGW	IWB70791	8	4948776	4.03	3608776-6288776	0.00	-0.11
Plant_height	IWB8326	6	510435	6.3266	0-1850435	0.12	-3.96
Plant_height	IWB7321	2	82475452	5.9572	81135452-83815452	0.07	12.95

Starting from the most associated SNP of the 3 main QTLs detected for yield components in VCU trial, the LD decay of 1.34 Mb was considered to detect the confidence interval under each QTL (Figure 18). Considering the confidence interval length for each QTL, the genetic interval was explored based on the *Triticum turgidum* cv Svevo *RefSeq* v1.0.
Table 10 reports the different candidate genes for each trait of interest. In general, the candidate genes are involved in secondary metabolism pathways, signal transduction with protein kinases and different transcription factor families. All the candidate genes for each QTL peak on each phenotypic traits were reported in Supplementary material (from Table S16 to Table S19).

Table 10: candidate genes from main GWAS peaks for the different agronomic traits: yield (corrected for 15% of moisture content), TGW, Plant height and heading date. The genes correspond to the most associated SNP marker from GWAS BLINK model. The gene functions were obtained from Ensembl Plants database.

Trait	Gene stable ID	Chr	Gene start (bp)	Gene end (bp)	Gene description
Yield_15	**TRITD6Bv1G132420**	**6B**	431951070	431989122	Prolyl oligopeptidase family protein
Yield_15	**TRITD7Bv1G198990**	**7B**	624077247	624078047	Germin-like protein
Yield_15	**TRITD5Av1G203790**	**5A**	546814761	546820749	F-box family protein
TGW	**TRITD5Av1G097200**	**5A**	279485390	279490053	Bifunctional uridylyltransferase/uridylyl-removing enzyme
TGW	**TRITD4Bv1G195230**	**4B**	646421171	646434994	Purple acid phosphatase
TGW	**TRITD2Av1G043500**	**2A**	94953246	94961150	Leucine--tRNA ligase
Plant height	**TRITD3Bv1G000220**	**3B**	508869	511463	Serine/threonine-protein kinase ATM G
Plant height	**TRITD1Bv1G030900**	**1B**	82473590	82475847	WRKY transcription factor
Heading date	**TRITD2Av1G019570**	**2A**	37397770	37408859	MLO-like protein
Heading date	**TRITD1Av1G004090**	**1A**	8966607	8967707	Glutathione S-transferase

In general, interesting genes can be identified in some agronomic traits based on their molecular function. For example, as regards yield, candidate genes are involved in metabolic pathways with the F-box family protein (*TRITD5Av1G203790*) being a strong candidate involved in different types of functions, among them there is the flowering time. Furthermore, other traits such as TGW and plant height have candidate genes connected with secondary metabolism and cell functions. As regards the heading date, it is worth mentioning that the QTL peak on chromosome 2A overlaps with *Ppd-2A* locus (Maccaferri et al., 2008). The most associated gene is the MLO-like protein (*TRITD2Av1G019570*), this gene family belongs to groups involved in stress response (Konishi et al., 2010) but it also have an influence on flowering time and inflorescence development. Considering the QTL confidence interval, in addition to the MLO-like protein, F-box transcription factors genes are reported, more connected with heading date and flower development (Table S19). Another interesting candidate is reported for plant height on chromosome 1B. The WRKY transcription factor (*TRITD1Bv1G030900*) represent a wide gene family of very common proteins in wheat that have very different functions and are involved in different metabolic pathways.
A more detailed analysis was performed for some of these candidate traits using Knetminer database. Starting from the most associated genes on *Triticum turgidum* cv Svevo assembly, *Triticum aestivum* cv Chinese Spring (CS) orthologues were obtained. The CS genes were used to explore the protein functions and gene network in which they are involved using the Knetminer database.

Candidate genes for yield, plant height and heading date were explored using Knetminer to better understand the gene network, used as main results to explore the gene network. As regards yield, CS orthologues were identified with the only exception of *TRITD5Av1G203790* (F-Box protein family) whose orthologue on chromosome 5A was not detected on *Triticum aestivum* cv Chinese Spring (CS), thus the 5B orthologue was used. The gene network highlights some principal traits in which the candidate genes for yield are involved, such as flowering time, seed size and protein content (Figure 22).

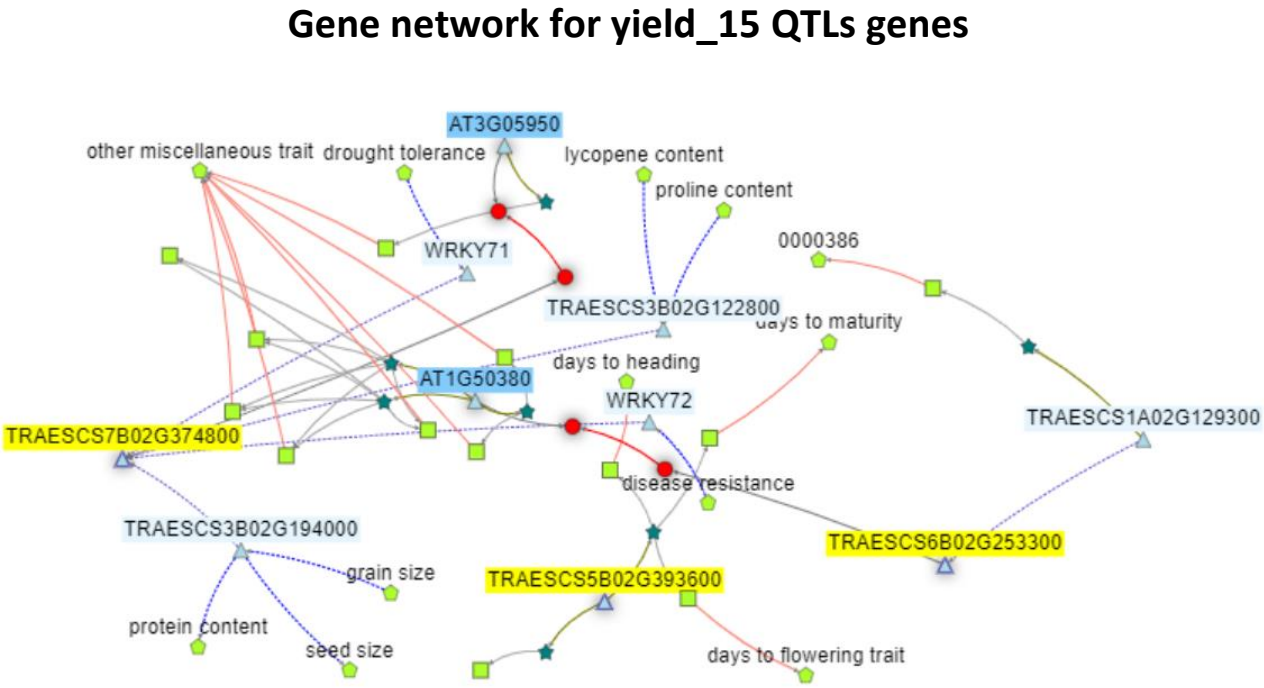


Figure 22: Knetminer gene function network from the most associated genes for yield trait. Gene functions are connected with flowering, protein content, days to maturity and seed size. The light green squares and pentagons represent the corresponding phenotype and trait of interest, the light blue triangles represent the genes involved in the network, the red circles report the proteins, the dark green stars represent the identified SNPs for the associated genes.

The main protein function, in which these genes are involved, corresponds to flowering time (CS homoeologues on chromosome 5B, corresponding to CS orthologue on chromosome 5A of *TRITD5Av1G203790*), protein content and grain size based on the gene networks (CS orthologue of the candidate gene *TRITD7Bv1G198990*) which can be directly connected with yield.
As regards plant height, the gene network, corresponding to candidate genes on chromosome 1B and 3B, shows specific molecular functions such as photomorphogenesis, pollen germination, cellular differentiation and organization (Figure 23). The main protein categories correspond to kinases and transcription factors, having molecular functions involved in plant development and differentiation.

Gene network for plant height QTLs genes

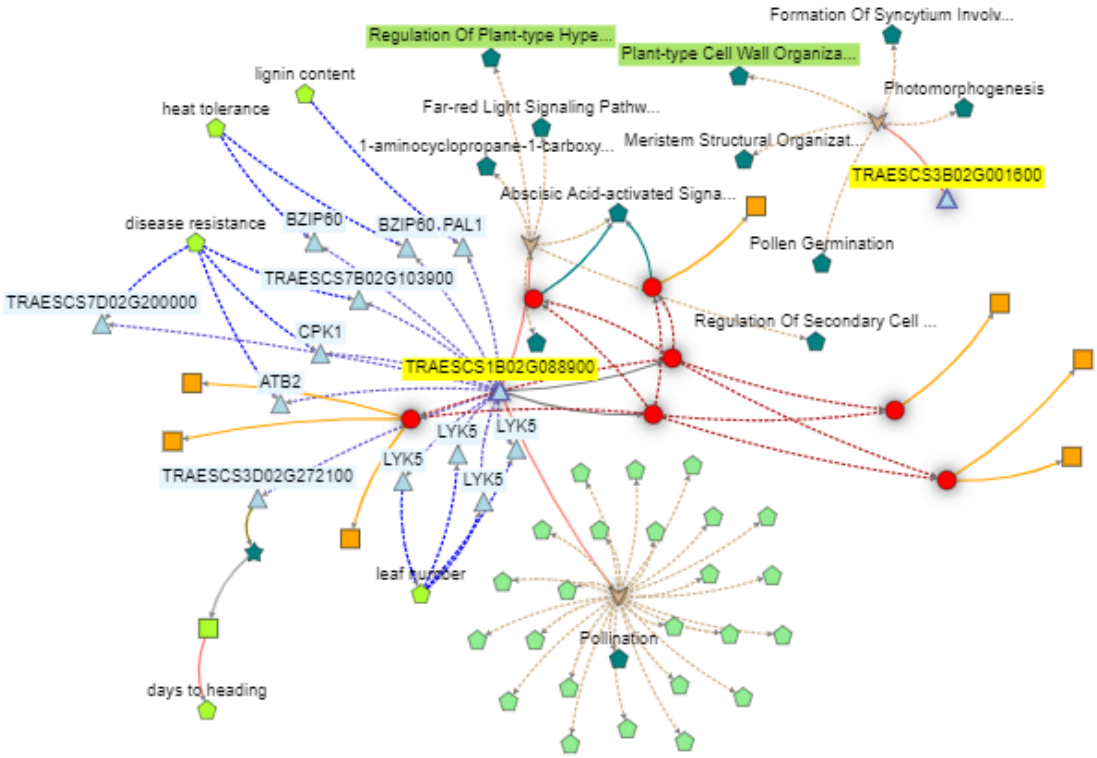


Figure 23: Knetminer gene network for plant height candidate genes. The main functions seem to be photomorphogenesis, heat tolerance, days to heading and cell development/differentiation. The light green squares and triangles represent the corresponding phenotype and trait of interest, the light blue triangles represent the genes involved in the network. The dark green pentagons report the biological process. The red circles report the proteins, the dark green stars represent the identified SNPs for the associated genes and the orange square report the reference pubblications on the NCBI database.

As regards heading date, gene functions of the *MLO15* genes on chromosome 2A have a strong connection with heading date and flowering time regulation. Moreover, the interval on chromosome 2A is next to and partially overlaps to the *Ppd-2A* QTL region (Maccaferri et al., 2008) (Figure 24). The *Ppd* locus is involved in controlling the flowering time and the photoperiod sensitiveness. In particular, the *Ppd-2A* locus has three alleles, the wild type confers sensitivity to the photoperiod and the other two alleles which are not sensible to it. The latter causes a very short pre-floweing phase in comparison to the sensitive alleles and its copy on the B genome (*Ppd-B1*) (Royo et al., 2016). It is worth pointing out that, beside the *MLO* gene seems to be involved also in flowering development, these gene belongs to a gene family principally involved to stress response (Konishi et al., 2010). In this QTL confidence interval, in addition to the *MLO15*, also F-box transcription factors genes are reported, more connected with heading date and flower development (Table S19).

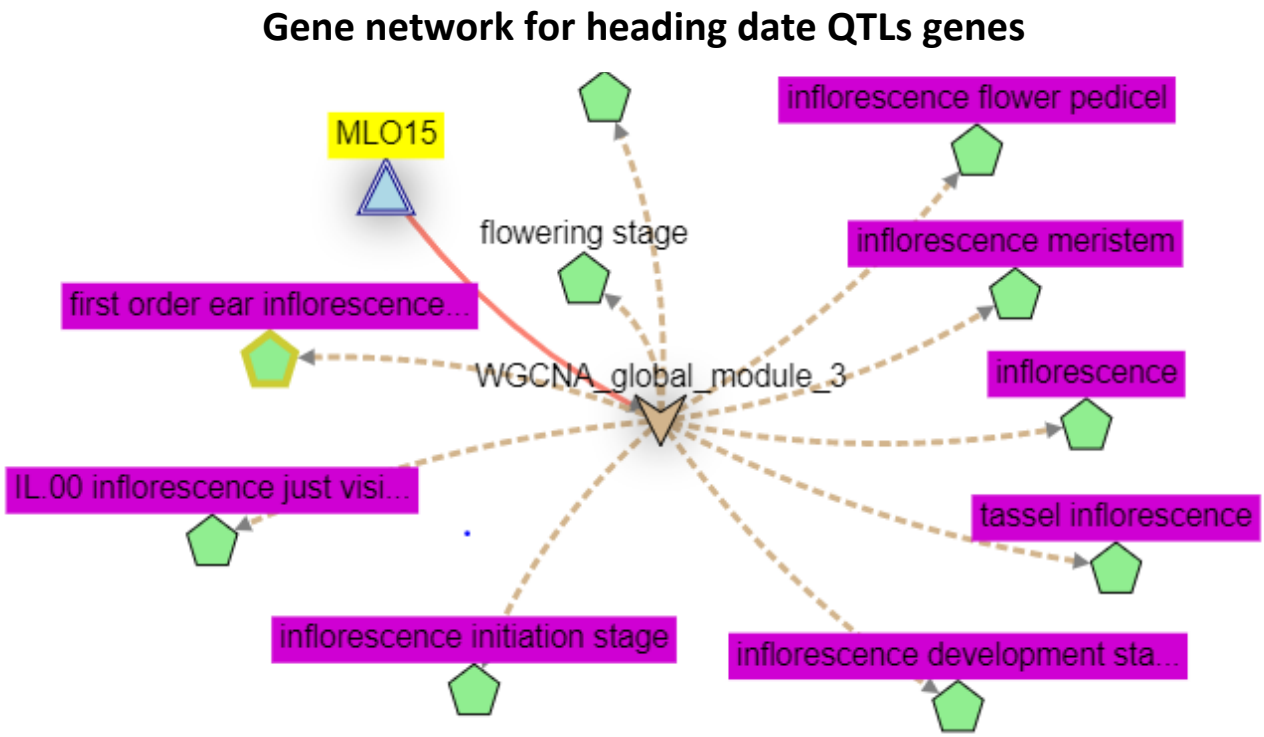


Figure 24: Knetminer gene network for heading date candidate genes. The main functions seem to be related to the inflorescence development and flowering stage. The light green pentagons represent the corresponding trait of interest, the light blue triangles represent the genes involved in the network. Brown arrows represent coexpression studies.

4.3.6 DUS GWAS results

As regards the DUS trial, the original hapmap including Innovar panel, as well as the GDP, durum panel 1 (DP1) and durum panel 2 (DP2), was filtered for the 232 DUS varieties sown in DUS trial (excluding CPVO controls). The LD decay and population structure results were the same as the ones already reported for the VCU trial (Figure 18). In fact, the different number of modern varieties included in the DUS trial comprehended all the VCU panel, thus these similar results were expected. The kinship matrix was computed with TASSEL5, and average genetic similarities were calculated (Figure 25).

Again, the most similar varieties were Natal, Ermes, Guadalso and Opera, which share similar breeding programs (CYMMIT, ICARDA, Southern Italian programs) and have a close genetic relationship with Avispa and Iride, previously detected as most genetically similar varieties in VCU panel. The more genetically distant varieties are RGT Voilur and Lupidur that, similar to the results detected for the VCU panel, come from northern European breeding programs very different from a major part of varieties included in DUS panel.

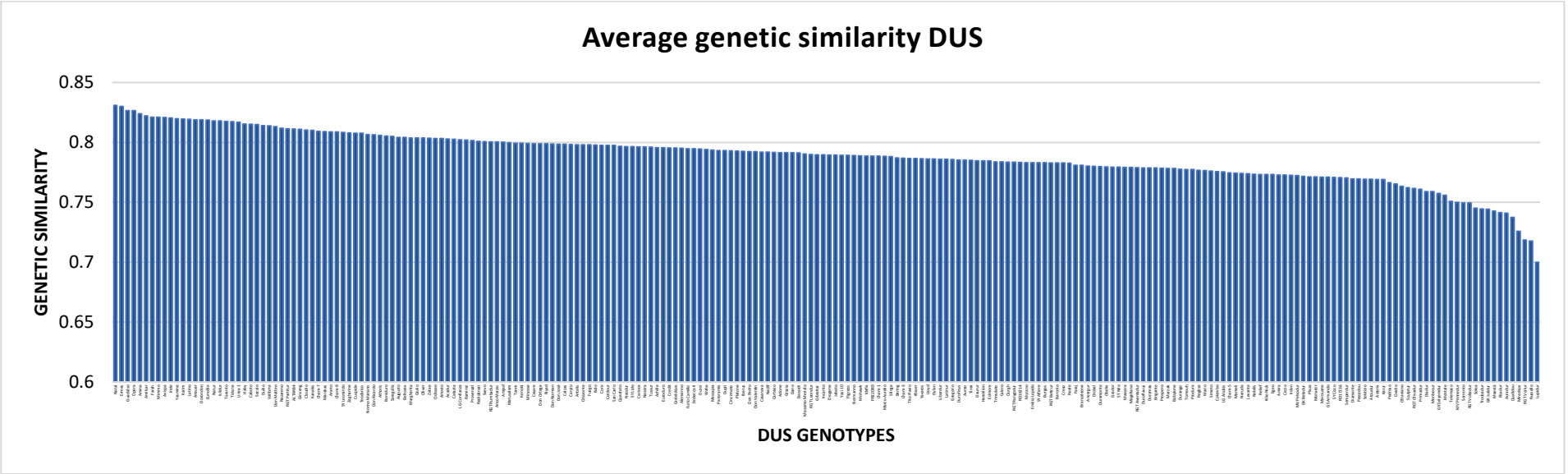
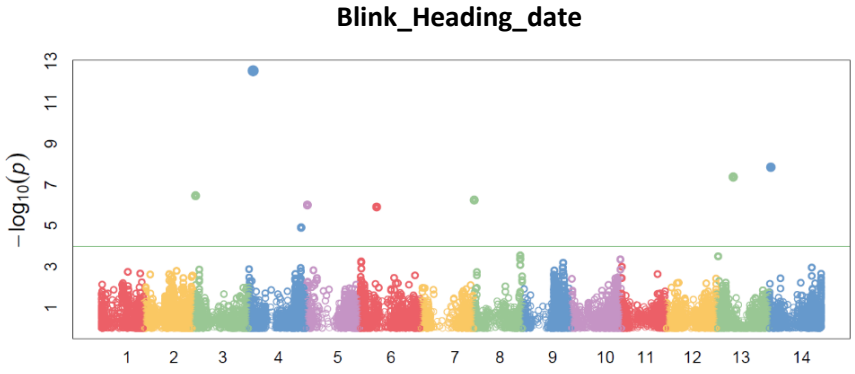
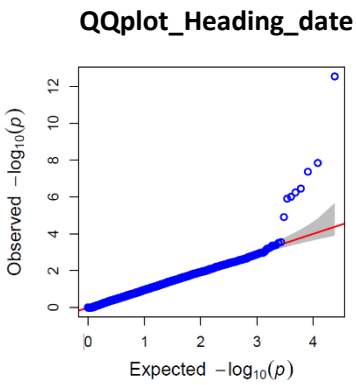
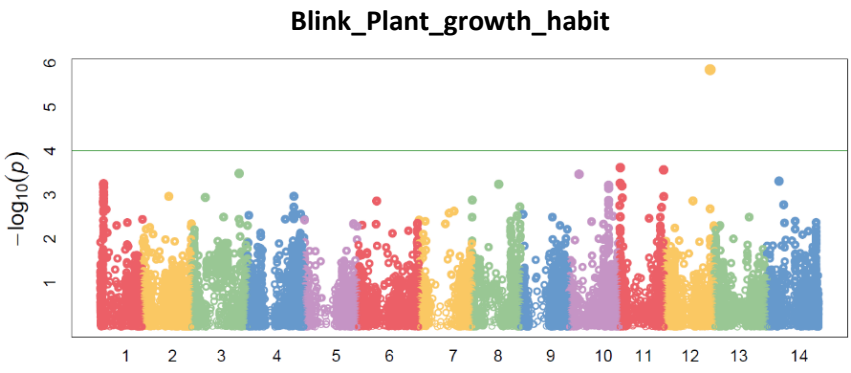


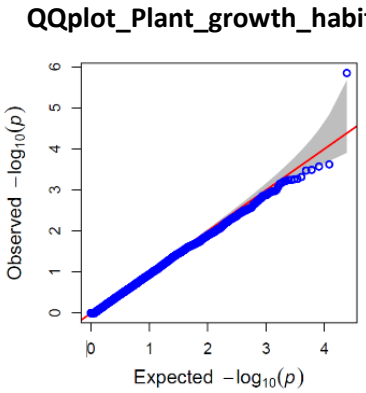
Figure 25: average genetic distance matrix based on kinship of DUS genotypic dataset. The average genetic distance was calculated for each variety and plotted in decrescent order. The x-axis reports the DUS genotypes and the y-axis report the value of genetic similarities calculated with identity by state.

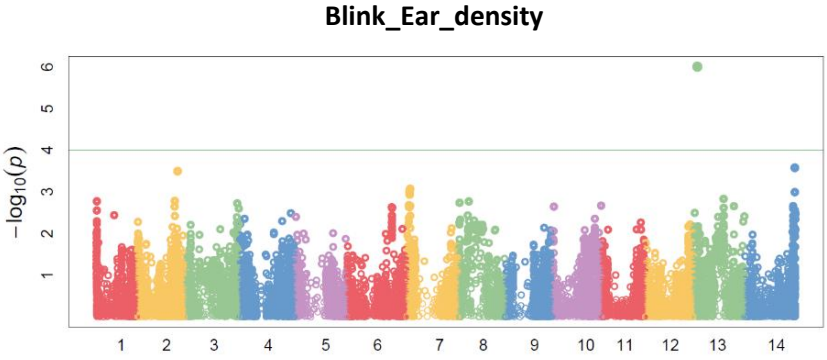
The GWAS analysis was performed like in the VCU panel, showing the results of Blink model corrected for population structure using kinship matrix. Differently from the VCU panel, the DUS trial is more focused on different morphological traits instead of yield, which aims at showing that a certain variety is substantially different from any other accession and that can receive the plant breeder's right (PBR). Herein are reported the principal results of GWAS analysis on DUS, with the main phenotypes that gave interesting QTL peaks. The GWAS Manhattan plots were performed on all the traits, with significant results, above all, on the following agronomic characters: heading date (calculated as number of days from the sowing date), plant growth habit, ear density (number of spikelets), ear glaucosity, culm glaucosity, awn colour and ear colour (Figure 26). All the previously reported GWAS models were tested, and results are reported in Supplementary materials from Figure S36 to Figure S42.

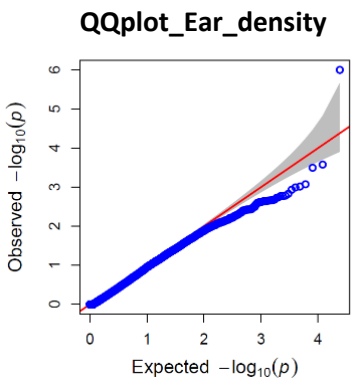












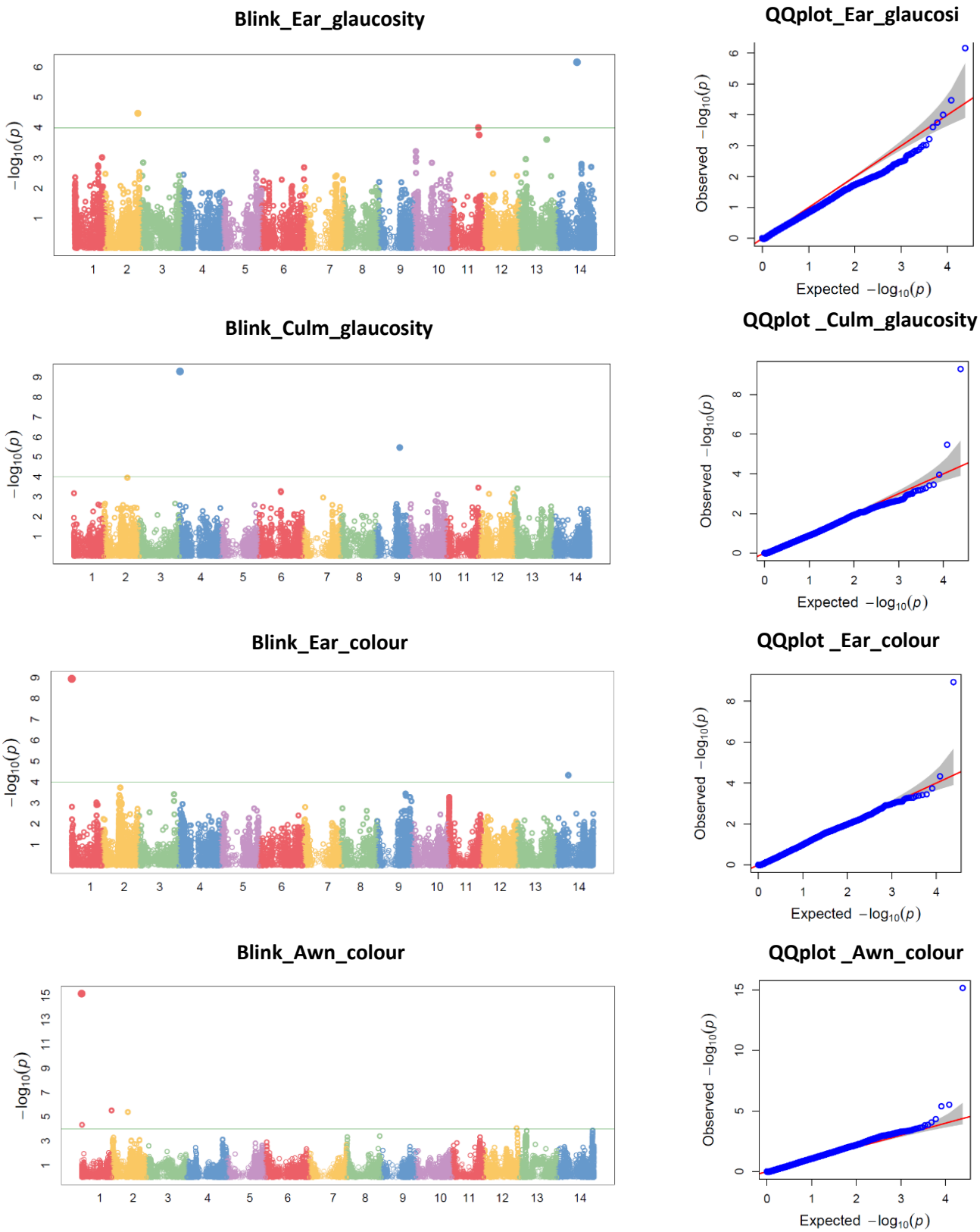


Figure 26: Manhattan plots and QQ plots from Blink for the main traits studied in DUS GWAS, such as heading date, growth habit, ear density, culm and ear glaucosity, ear colour and awn colour.

Based on the GWAS plots, the summary of the QTL peaks shows the most associated peaks which are reported for each trait (Table 11). As regards candidate gene analysis, the underlined peaks in Table 11 were considered.

Table 11: most associated SNP based on GWAS results for each trait are reported on DUS dataset. The main peaks considered from major QTLs are underlined. For each trait the confidence interval (C. I.), the R^2 linear effect and the effect of the minor allele on the major allele based on the GLM model are reported.

SNP	Chr	Position	-logP	C.I (+/-LD decay)	R2_GLM	Effect_GLM	Trait
IWB70422	4	56659272	12.6	55319272-57999272	0.28	1.94	Heading date
IWB13248	14	22091813	7.85	20751813-23431813	0.05	-3.07	Heading date
IWB13689	13	230040105	7.37	228700105-231380105	0.10	2.14	Heading date
IWA2526	3	36293364	6.46	34953364-37633364	0.18	-1.88	Heading date
IWB74226	8	4947809	6.25	3607809-6287809	0.23	1.82	Heading date
IWB2295	5	17213309	6.01	15873309-18553309	0.04	-0.91	Heading date
IWB31829	6	228908358	5.91	227568358-230248358	0.03	-1.59	Heading date
IWB36729	12	638496572	5.85	637156572-639836572	0.11	0.30	Plant growth habit
IWB55999	13	49529009	6	48189009-50869009	0.08	0.26	Ear density
IWB4352	14	389033953	6.16	387693953-390373953	0.07	-1.65	Ear glaucosity
IWB3330	2	625628003	4.47	624288003-626968003	0.10	0.48	Ear glaucosity
IWA4928	11	538824385	4	537484385-540164385	0.10	0.46	Ear glaucosity
IWB51601	4	9533722	9.28	8193722-10873722	0.17	0.27	Culm glaucosity of the neck
IWB30537	9	465823635	5.47	464483635-467163635	0.07	0.29	Culm glaucosity of the neck
IWB525	1	1104522	8.93	0-2444522	0.13	-0.20	Ear colour
IWB59968	14	243588961	4.33	242248961-244928961	0.04	-0.24	Ear colour
IWB72726	1	1161799	15.2	0-2501799	0.37	-0.54	Awn colors
IWB52266	1	582448730	5.52	581108730-583788730	0.08	0.45	Awn colors
IWB23218	2	314321268	5.39	312981268-315661268	0.01	0.35	Awn colors
IWB46412	1	9220037	4.34	7880037-10560037	0.08	0.24	Awn colors
IWB71618	12	641564818	4.06	640224818-642904818	0.14	-0.37	Awn colors

For all the studied traits, the main QTLs were considered to explore candidate genes within the confidence interval (underlined peaks in Table 11). This interval was again calculated from the LD decay of 1.34 Mb (Figure 18). The complete gene intervals for all the major QTLs were reported in Supplementary materials from Table S20 to Table S25. The candidate genes are reported in Table 12, obtaining the gene function from Ensembl Plants database. The genes correspond to the closest genes to the most associated SNP markers from GWAS analysis.

Table 12: candidate genes for DUS traits analyzed with GWAS analysis using BLINK model. The candidate genes are reported for heading date, plant growth habit, ear density, ear and culm glaucosity, ear and awn colour. Gene functions were retrieved from Ensembl Plant database.

Traits	Gene stable ID	Chr	Gene start (bp)	Gene end (bp)	Gene description
Heading date	TRITD2Bv1G025540	2B	56658223	56660088	Serine/threonine-protein kinase ATM
Heading date	TRITD7Bv1G008490	7B	22089866	22091774	Late embryogenesis abundant protein
Heading date	TRITD7Bv1G008520	7B	22093924	22094451	F-box-like protein
Heading date	TRITD7Av1G095190	7A	230038058	230039686	(RAP Annotation release2) Galactose-binding like domain containing protein
Heading date	TRITD7Av1G095200	7A	230040373	230047509	Glycine-rich domain-containing protein 2 G
Heading date	TRITD2Av1G019050	2A	36289642	36293816	Kinase interacting (KIP1-like) family protein
Plant growth habit	TRITD6Bv1G205280	6B	637988454	637992747	Autophagy-related protein 18
Plant growth habit	TRITD6Bv1G205440	6B	638501886	638502842	Tyrosine-protein phosphatase CpsB
Ear density	TRITD7Av1G025250	7A	49243760	49244887	1-aminocyclopropane-1-carboxylate oxidase homolog 2
Ear density	TRITD7Av1G025340	7A	49929392	49930359	Auxin repressed/dormancy associated protein
Ear glaucosity	TRITD7Bv1G123690	7B	389032006	389033986	40S ribosomal protein S13
Ear glaucosity	TRITD1Bv1G205170	1B	625621594	625625190	F-box family protein
Ear glaucosity	TRITD1Bv1G205230	1B	625739966	625742854	Glycosyltransferase
Culm glaucosity of the neck	TRITD2Bv1G004670	2B	9533137	9535194	Cytochrome P450
Ear colour	TRITD1Av1G000450	1A	1095301	1106930	Paired amphipathic helix protein Sin3
Awn colour	TRITD1Av1G000480	1A	1156772	1162539	Phospholipid-transporting ATPase

The main genes for DUS related traits corresponded to metabolic pathways, transcription factors and signal transduction proteins. To better survey the gene functions, as already performed for VCU candidate genes, orthologues were obtained from CS genome and gene networks were enquired using Knetminer just for some of the studied traits, such as: heading date, plant growth habit, ear density, ear glaucosity and ear colour.

As regards heading date, candidate genes seem to be involved in inflorescence development, photoperiod, and flowering date (Figure 27). Furthermore, the interval on chromosome 2A is the same already detected in VCU panel very close to the *Ppd-2A* QTL, whose B copy (*Ppd-2B*) was detected also in the DUS matching perfectly the QTL peak on chromosome 2B. However, the most associated gene on the 2A chromosome was different from the one reported for VCU trials, although the confidence interval is the same both for VCU and DUS panels heading date.

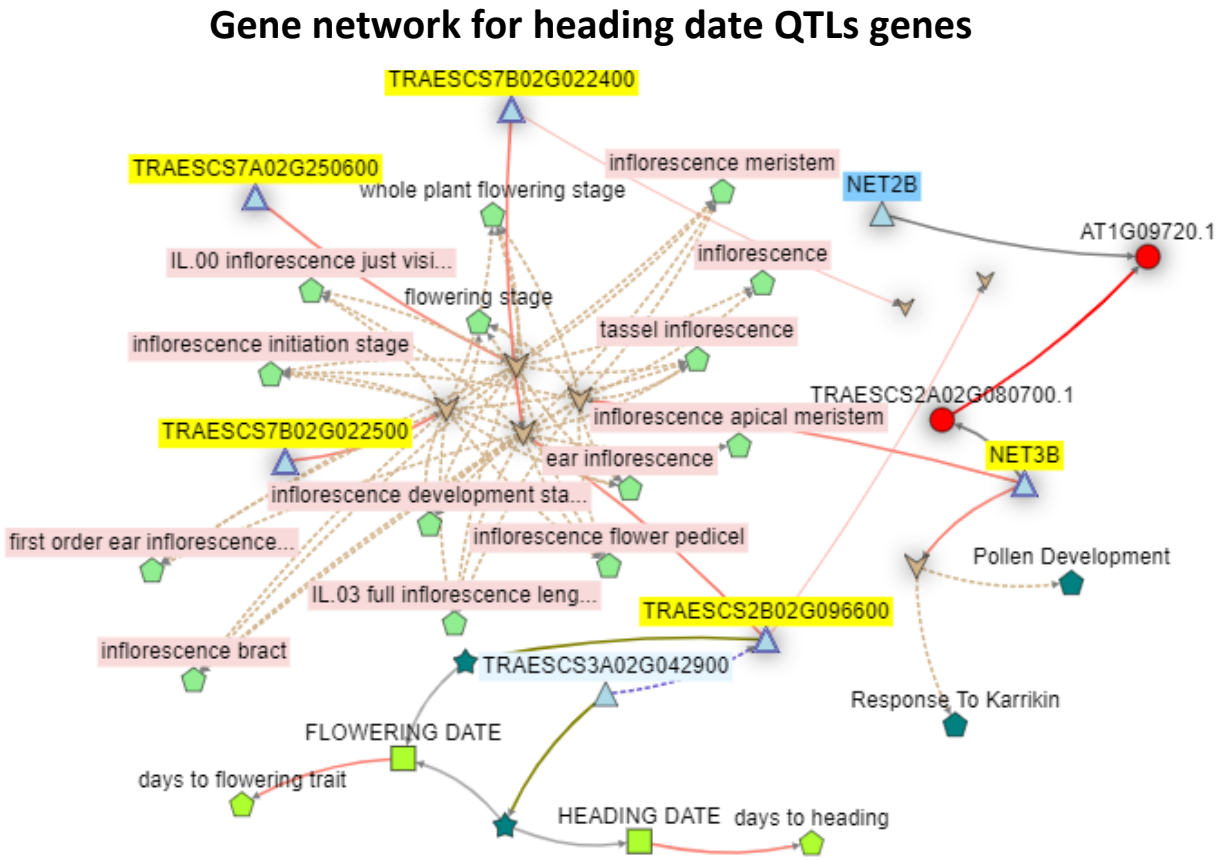


Figure 27: Knetminer gene network for heading date. Principal gene functions where genes are involved are flowering time, inflorescence development. The light green squares represent the corresponding phenotype, the light blue triangles represent the genes involved in the network. The red circles represent the proteins. The dark green pentagons correspond to molecular processes and brown arrows represent coexpression studies.

The gene network for plant growth habit trait was explored, the orthologue on CS of *TRITD6Bv1G205280* had connections with two transcription factors (Figure 28). The *LBD20* gene relates to lateral roots development, and it belongs to transcription factor gene family involved in jasmonate pathway. The *WRKY28* genes belongs to a transcription factor family with several different functions in wheat. These genes in *Arabidopsis thaliana* seem to be involved in salicylic acid pathway, regulating plant development to abiotic stress acting also on roots as expression tissue.

Gene network for plant growth habit QTLs genes

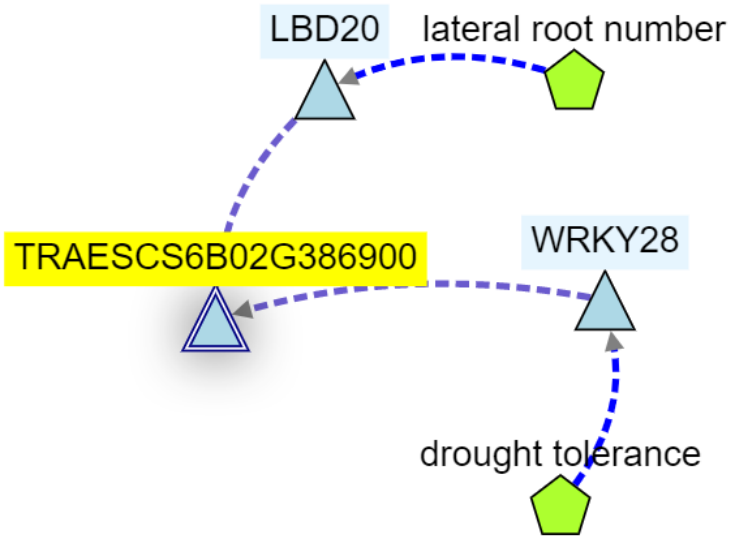


Figure 28: Knetminer gene network of candidate genes for plant growth habit. The candidate genes seem to be involved in root traits and drought tolerance. The light green pentagons represent the corresponding trait of interest, the light blue triangles represent the genes involved in the network.

Also, the genes associated with ear density seem to be involved in inflorescence architecture, spike development at different growing stages (Figure 29). Also in this case, no orthologues were found on chromosome 7A in CS for both genes, only the gene on chromosome 7D on CS for *TRITD7Av1G025340*.

Gene network for ear density QTLs genes

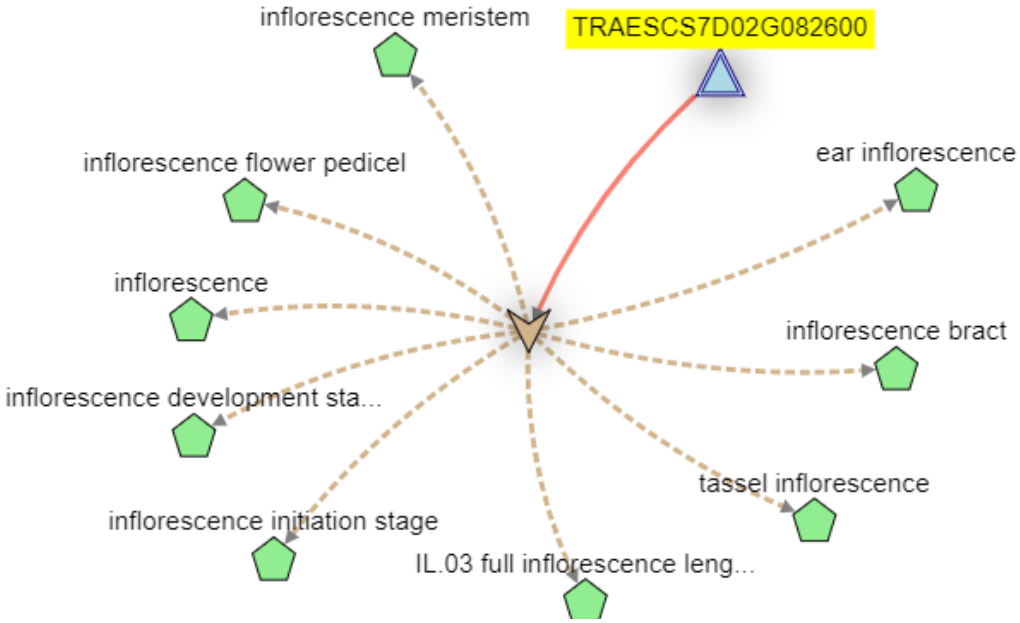


Figure 29: knetminer genes functions for ear density traits. All the genes are involved in inflorescent development pathways as general functions. The light green squares represent the corresponding phenotype, the light blue triangles represent the genes involved in the network. Brown arrows represent coexpression studies.

As regards other spike traits, the functions of genes associated to ear glaucosity are strongly linked to ear inflorescence. The functions of these genes are like the ones already detected for other traits linked with spike inflorescence. General functions are ascribed to development and structure of ear floret. On the other hand, *RPS13A* (*TRITD7Bv1G123690*) seemed to be involved, in *Arabidopsis thaliana*, on leaf morphogenesis and trichome development (Figure 30).

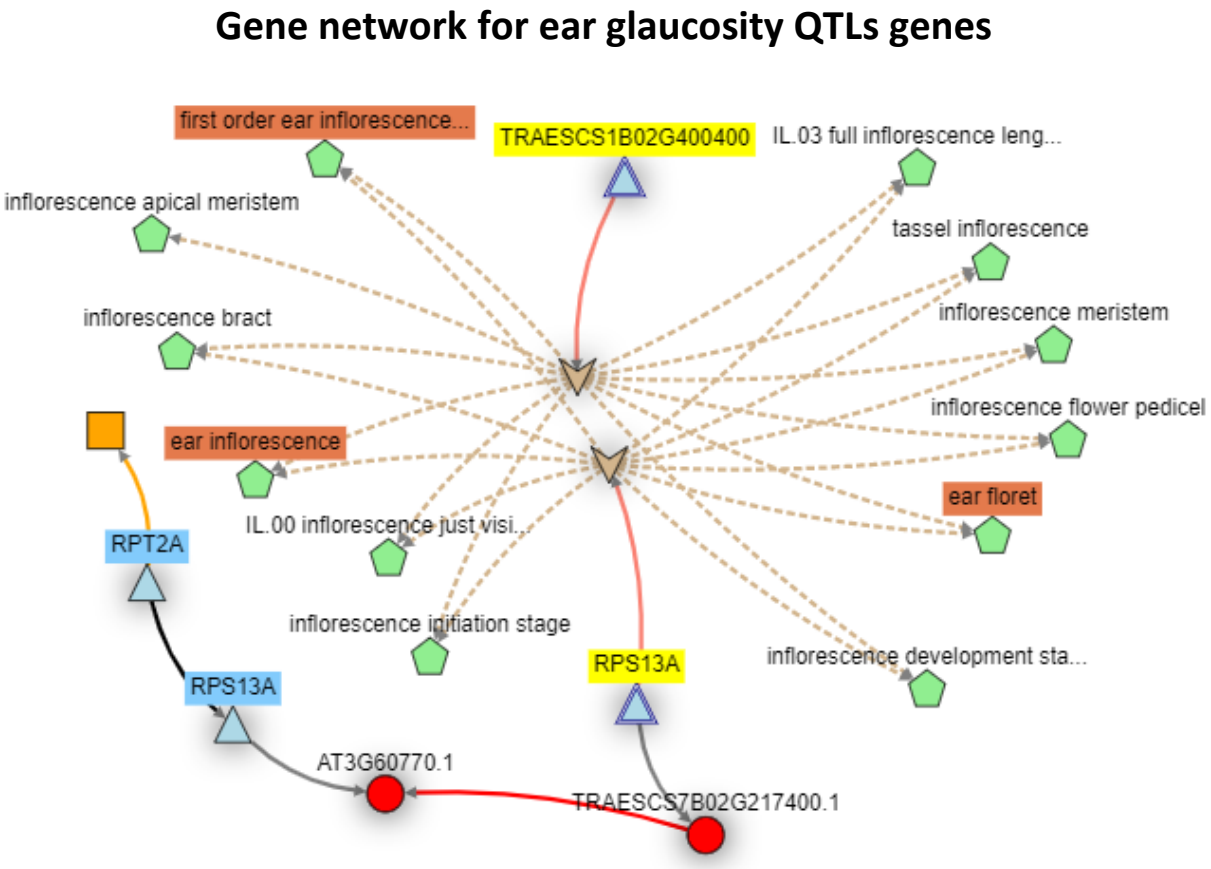


Figure 30: knetminer gene functions for most associated genes on ear glaucosity trait. In general, the functions are involved in ear inflorescence and development at different growth stages. The RPS13A gene is involved in thricome development. The light green pentagons represent the corresponding traits of interest, the light blue triangles represent the genes involved in the network. Brown arrows represent coexpression studies and the red circles report the associated proteins.

The last spike trait regards ear colour which share the same peak on chromosome 1A. The CS orthologue is gene *SNL5* (*TRITD1Av1G000450*) which is linked to *HD1* gene involved in flowering development in rice and other species, and gene *SNL5* interacts with *ERF* gene repressing genes in abscisic acid and drought stress response. The generic functions correspond to ear floret and inflorescence development. The *SNL5* gene, in *Arabidopsis thaliana*, is involved in regulation of transcriptional activity and it is expressed in inflorescence, stem, leaves (Figure 31).

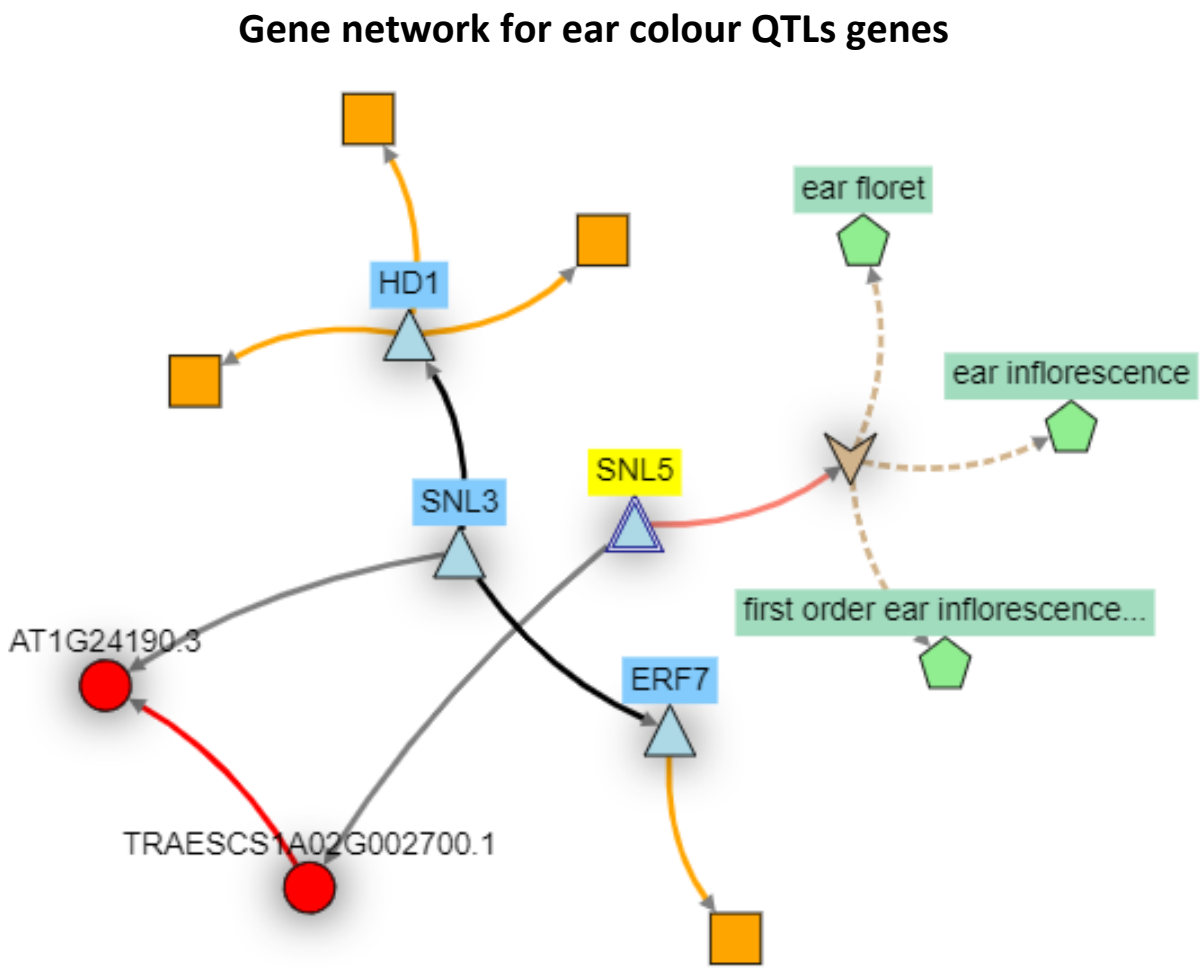


Figure 31: knetminer function of candidate genes associated with ear colour. Genes SNL3, SNL5 and HD1 are involved in ear development, transcription factors and infloresence architecture. The red circles represent protein. The light red pentagons represent traits of interest in which the gene network is involved and the orange squares the publications associated with the phenotypic traits based on the NCBI database.

4.4 Discussion

The VCU and DUS Innovar durum trials were carried out in different environments and phenotypes with EU CPVO protocols during the growing season. The VCU trials were more connected with yield related traits, as the disease pressure during the growing season was not relevant enough to proper analyse the data. On the other hand, the DUS trials are more connected with morphological traits related to plant development, colour and glaucosity of ears, culms and flag leaves, and shape of leaves or ear components. The DUS and VCU trials reflect the process of variety registration nowadays used in EU CPVO to register a genotype as new variety in the national and European catalogues. A variety can be then commercialized in the European market.

The aim of these protocols is to register newly bred varieties demonstrating the distinctiveness, uniformity and stability (DUS) together with the value of cultivation and use (VCU) to demonstrate that newly bred varieties are different from the already registered ones.

The principal aim of the PhD thesis on Innovar project is to augment the necessary information for VCU and DUS protocols with molecular data from SNPs markers, QTLs or candidate genes responsible for phenotypic variation of yield and morphological traits studied in the VCU and DUS trials.

Some of the QTLs identified in VCU and DUS analysis, corresponds to other cloned QTLs responsible for similar traits, considering the major identified QTLs.

As regards VCU dataset, the yield_15 major QTL interval on chromosome 6B was explored considering the population LD decay to identify the confidence interval from the most associated SNP IWB49077, from 430,648,767 bp to 433,328,767 bp (2.68 Mb).

The 6B interval corresponds to other QTLs detected, such as a grain protein contend (GPC) QTLs different chromosomes which are related to high values of GPC in durum wheat Canadian lines (maximum values detected in Strongfield variety) (Suprayogi et al., 2009). Furthermore, Graziani et al. (Graziani et al., 2014) detected different QTLs connected with yield related traits on the RIL population Kofa x Svevo, identifying a QTL on chr6B related to test weight in durum wheat. Other minor QTLs on the chr6B region were identified for kernel weight (Roncallo et al., 2017).

As for heading date, the QTL on chromosome 2A was considered, with the most associated marker IWB70098 and the confidence interval between 36,057,799 bp and 38,737,799 bp. Different QTLs were detected for other traits correlated with heading date, such as number of kernels/m^2 (Graziani et al., 2014; Roncallo et al., 2017) and for plant height and grain yield (Milner et al., 2016). However, the most associated peak identified on the VCU panel for heading date does not perfectly overlaps with the known *Ppd-2A* locus (Maccaferri et al., 2008) and it is very close to it, but based on the gene network analysis it seems anyway involved in flowering time regulation.

TGW confidence interval was analysed from the major QTL on chromosome 5A, 278,445,716 bp and 281,125,716 bp. Related QTLs were detected on spike harvest index (Peleg et al., 2009; Tzarfati et al., 2014), spike fertility for the number of grains per spike and spikelet (Giunta et al., 2018) and other QTL for heading date (Maccaferri et al., 2011c).

For plant height, the most associated SNP was IWB8326 in the proximal part of chromosome 5B with a confidence interval from the start to 1,850,435. Only one QTL was detected in these traits from literature, involved in grain yield and TGW in durum wheat (Roncallo et al., 2018).

As regards the DUS, 7 morphological traits were taken into consideration and the main QTLs were considered for the candidate gene evaluation. As for heading date, the main QTL was on chromosome 2B with the most associated peak IWB70422 between 55,319,272 bp and 57,999,272 bp. Different QTLs were identified for spike grains and architecture (Milner et al., 2016; Patil et al., 2013; Roncallo et al., 2017), grain protein content (Marcotuli et al., 2017) and *Ppd-2B* involved in

photoperiod regulation (Maccaferri et al., 2008). Others are associated with flowering time and development, such as days to anthesis (Giunta et al., 2018). The *Ppd* locus is associated with photoperiod sensitiveness, thus with flowering time. The homoeologous gene on the A genome has two alleles associated with short flowering period, on the other hand the B gene is, instead, associated with sensibility to photoperiod and flowering time (Royo et al., 2016). The GWAS analysis on VCU heading date trait resulted in the identification of a QTL peak very close to the *Ppd-2A*, on the other hand the GWAS on DUS dataset showed a major QTL on the B copy (*Ppd-2B)* and a secondary QTL on chromosome 2A on the same region already identified in the VCU core trial, in addition to other QTLs. This discrepancy between GWAS on DUS and VCU results on heading date may be caused by the different genotypes included in the two datasets. In fact, the VCU panel is composed of durum wheat varieties mainly adapted to a Mediterranean environment. As a consequence, the major QTL close to the *Ppd-2A* was detected in the VCU panel, associated also with alleles unsensitive to photoperiod and with early flowering time. On the other hand, the *Ppd-2B* copy is sensitive to the photoperiod, the DUS panel is composed of different genotypes adapted to single environments, coming also from northern European breeding programs (Figure 11), which are characterized by late flowering time.

As regards plant growth habit analysed in DUS trials, the major QTL was on chromosome 6B with the most associated marker IWB36729 and a confidence interval between 63,7156,572 bp and 639,836,572 bp. Based on literature research, interesting QTLs peaks are connected with final yield and probably are closely related to the elevation angle and growth habit of the accessions. Mengistu et al. (Mengistu et al., 2016) showed that this region is also connected with QTL associated with number of spikes per plant. This is directly associated also with the number of tillers per plant, identified in a specific QTL by Giunta et al. (Giunta et al., 2018).

The last character associated with yield in DUS GWAS was the ear density, identified as the number of spikelets per spike. The major QTL is on chromosome 7A, with the most associated SNP IWB55999 and a confidence interval between 48,189,009 bp and 50,869,009 bp. As detected also for other traits, the region is the same of QTLs associated with yield component and grain protein content (Blanco et al., 2012).

As regards glaucosity, DUS traits monitor both ear and culm plant tissues. The ear glaucosity has a major QTL on chromosome 2B, with marker IWB4352 as the most associated identifying a region between 387,693,953 bp and 389,033,953 bp. The QTL associated with ear galucosity was confirmed also by the detection of a QTL involved in glabrousness in that region (Distefield et al., n.d.).

Ear and awn colour share the same peak proximal region of chromosome 1A, with a very strong -logP value. The ear colour on 1A is associated with the most associated SNP IWB525 with the confidence interval that goes from the chromosome start to 2,444,522 bp. In the same region Patil et al. (2008) published a QTL responsible for yellow pigment content, together with other QTLs on chromosomes 3B, 5B, 7A and 7B related to the same trait. As regards the awn colour, as the region is the same the same QTLs from literature were identified, although the most associated SNP is IWB72726.

The GWAS analysis on VCU and DUS durum datasets resulted in several interesting peaks for all the traits. The reported work focuses on yield related traits for VCU dataset, and on the main morphological traits results for the DUS dataset. However, different QTLs were detected for all the traits phenotypes in the protocols (not all reported in the work). As for the yield related traits, the candidate genes seem to be involved with grain protein content, inflorescence architecture and development, flowering time, kernel numbers and weight. Some of these functions are also conserved in the DUS trait candidate genes, but the morphology and ear colour associated genes seem to be mainly involved in general secondary metabolism pathways. The consistency of the peaks is also confirmed by different QTLs detected in literature for the same traits.

The procedure followed in the Innovar project gives a complete overview of differences and characteristics of durum wheat commercial varieties, both from a phenotypic and a genotypic point of view. The genotyping and population structure approaches allowed to identify genetic similarities and ancestry relationships between genotypes, identifying subpopulations and genotypic groups corresponding to specific breeding programs and origins. The VCU and DUS phenotyping data were analyzed for multi-enviromental trials, this approach allows to detect GxE relationship for each variety to the different agro-climatic zones (ACZ), correcting phenotypic data for the field trials conditions. Different phenotypic data from VCU and DUS analysis were used for the GWAS analysis using different statistical models, this approach allowed to identify the main QTLs for each trait confirmed across the different models assuring high confidence for gene interval exploration.

The genotyping work performed on varieties included in the DUS panel allows to compare different commercial varieties based on genetic distances. This comparison may allow to identify the rate of genetic and morphological similarity (due to common breeding origins and ancestors) of varieties under registration procedure in comparison to already commercialized varieties, thus identifying similarity thresholds to use as cutoff for new varieties during the registration process. Moreover, the DUS protocols are only based on morphological characters, the identification of molecular markers associated to these characters can augment the precision and information useful to better discriminate varieties in the registration procedures. The use of molecular markers associated with important phenotypic traits is particularly useful also in a VCU perspective, as the protocols deals mainly with yield related phenotypic traits and disease resistance. The identification of molecular markers, or haplotypes, associated to resistance to pathogens or increase yield productions may play an important role in the identification of better performing varieties useful for a more sustainable agriculture in a climate change perspective. In particular, markers associated main QTLs for each trait can be converted in KASP markers, being then used for genetic improvement and genotype selection during variety registration procedure as well as for research purposes.

This result facilitates breeder's work for variety selection and MAS, as well as it could represent an implementation of the DUS and VCU protocols facilitating the variety registration steps. These protocols could be no more based only on morphological traits, but a genetic connection will be established with the major impact QTL. The results of Innovar projects should be disseminated at EU and CPVO levels, which is starting to consider molecular markers as important resources for variety registration protocols (Yu and Chung, 2021).

4.5 Supplementary material

Table S13: List of durum wheat varieties sown in VCU trials. The specific research centre or institution of is reported for each variety, specifying the field trial location where the VCU trial occurred.

Name	Type
Antalis	Control
Claudio	Control
Don Ricardo	Control
Iride	Control
Monastir	Control
Simeto	ACZ Maritime South
Marco Aurelio	ACZ Maritime South
Saragolla	ACZ Maritime South
Svevo	ACZ Maritime South
Odisseo	ACZ Maritime South
Amilcar	ACZ Mediterranean
Don Norman	ACZ Mediterranean
Don Ortega	ACZ Mediterranean
Euroduro	ACZ Mediterranean
Sculptur	ACZ Mediterranean
Gk Betadur	ACZ Panonian
Gk Julidur	ACZ Panonian
Gk Selyemdur	ACZ Panonian
Mv Pelsodur	ACZ Panonian
Mv Pennedur	ACZ Panonian
Aneto	Site Specific-CSIC
Calcas	Site Specific-CSIC
Calero	Site Specific-CSIC
Carpio	Site Specific-CSIC
Duroi	Site Specific-CSIC
Emilio Lepido	Site Specific-CSIC
Ottaviano	Site Specific-CSIC
Prospero	Site Specific-CSIC
Qualidou	Site Specific-CSIC
Solea	Site Specific-CSIC
Taranto	Site Specific-CSIC
Trimulato	Site Specific-CSIC
Ovidio	Site Specific-CSIC & HORTA & SIDI & Marchouch
Fuego	Site Specific-CSIC & UNIDEB
Teodorico	Site Specific-CSIC & UPM
Daurur	Site Specific-CSIC & UPM &UNITUS & HORTA & UNIDEB &SIDI &Marchouch
Ramirez	Site Specific-CSIC & UPM &UNITUS & HORTA & UNIDEB &SIDI &Marchouch
Achille	Site Specific-HORTA
Ariosto	Site Specific-HORTA
Brigante	Site Specific-HORTA
Ciccio	Site Specific-HORTA

Diogene	Site Specific-HORTA
Don Matteo	Site Specific-HORTA
Duilio	Site Specific-HORTA
Kanakis	Site Specific-HORTA
Linea 1	Site Specific-HORTA
Linea 2	Site Specific-HORTA
Maesta	Site Specific-HORTA
Martinur	Site Specific-HORTA
Minosse	Site Specific-HORTA
Nadif	Site Specific-HORTA
Natal	Site Specific-HORTA
Nonno Mariano	Site Specific-HORTA
Normanno	Site Specific-HORTA
Pigreco	Site Specific-HORTA
PR22D89	Site Specific-HORTA
RGT Aventadur	Site Specific-HORTA
RGT Voltadur	Site Specific-HORTA
San Carlo	Site Specific-HORTA
Telemaco	Site Specific-HORTA
Anco Marzio	Site Specific-HORTA & SIDI & Marchouch
Aureo	Site Specific-HORTA & SIDI & Marchouch
RGT Leondur	Site Specific-HORTA & SIDI & Marchouch
Sfinge	Site Specific-HORTA & SIDI & Marchouch
Zetae	Site Specific-HORTA & SIDI & Marchouch
Amjad	Site Specific-SIDI & Marchouch
Beltorax	Site Specific-SIDI & Marchouch
Cham1	Site Specific-SIDI & Marchouch
Cham3	Site Specific-SIDI & Marchouch
Cham5	Site Specific-SIDI & Marchouch
Cham7	Site Specific-SIDI & Marchouch
Cham9	Site Specific-SIDI & Marchouch
Core	Site Specific-SIDI & Marchouch
Egeo	Site Specific-SIDI & Marchouch
Faraj	Site Specific-SIDI & Marchouch
Gibraltar	Site Specific-SIDI & Marchouch
Irden	Site Specific-SIDI & Marchouch
Karim	Site Specific-SIDI & Marchouch
Margherita	Site Specific-SIDI & Marchouch
Marouane	Site Specific-SIDI & Marchouch
Nassira	Site Specific-SIDI & Marchouch
SY Nilo	Site Specific-SIDI & Marchouch
Tarek	Site Specific-SIDI & Marchouch
Zagharin	Site Specific-SIDI & Marchouch
Atoudur	Site Specific-UNIDEB
Augusto	Site Specific-UNIDEB
Auradur	Site Specific-UNIDEB
Babylone	Site Specific-UNIDEB

Casteldoux	Site Specific-UNIDEB
Diadur	Site Specific-UNIDEB
Duramonte	Site Specific-UNIDEB
Durofinus	Site Specific-UNIDEB
Duroflavus	Site Specific-UNIDEB
Duromax	Site Specific-UNIDEB
Elsadur	Site Specific-UNIDEB
Karur	Site Specific-UNIDEB
Lupidur	Site Specific-UNIDEB
Mario	Site Specific-UNIDEB
Monbecur	Site Specific-UNIDEB
Nobilis	Site Specific-UNIDEB
Panoramix	Site Specific-UNIDEB
Rangodur	Site Specific-UNIDEB
RD16214	Site Specific-UNIDEB
RD17236	Site Specific-UNIDEB
Rgt Voilur	Site Specific-UNIDEB
Troubadur	Site Specific-UNIDEB
Anubis	Site Specific-UNIDEB & SIDI & Marchouch
Alfiere	Site Specific-UNITUS
Bering	Site Specific-UNITUS
Caboto	Site Specific-UNITUS
Cartesio	Site Specific-UNITUS
Casanova	Site Specific-UNITUS
Colombo	Site Specific-UNITUS
Dylan	Site Specific-UNITUS
Federico II	Site Specific-UNITUS
Furio Camillo	Site Specific-UNITUS
Gregorio	Site Specific-UNITUS
Heraklion	Site Specific-UNITUS
Idefix	Site Specific-UNITUS
Latinur	Site Specific-UNITUS
Lloyd	Site Specific-UNITUS
Mameli	Site Specific-UNITUS
Nazareno	Site Specific-UNITUS
Orobel	Site Specific-UNITUS
Platone	Site Specific-UNITUS
Quadrato	Site Specific-UNITUS
SY Leonardo	Site Specific-UNITUS
Tiziana	Site Specific-UNITUS
Montale	Site Specific-UNITUS & HORTA
Redidenari	Site Specific-UNITUS & HORTA & SIDI & Marchouch
Levante	Site Specific-UNITUS & HORTA & UNIDEB
Secolo	Site Specific-UNITUS & HORTA & UNIDEB & SIDI & Marchouch
Cesare	Site Specific-UNITUS & SIDI & Marchouch
Ermes	Site Specific-UNITUS & SIDI & Marchouch
Marakas	Site Specific-UNITUS & SIDI & Marchouch

Anvergur	Site Specific-UNITUS & UNIDEB
Miradoux	Site Specific-UNITUS & UNIDEB
Obelix	Site Specific-UNITUS & UNIDEB
Relief	Site Specific-UNITUS & UNIDEB
Avispa	Site Specific-UPM
Bolo	Site Specific-UPM
Boniduro	Site Specific-UPM
Catasta	Site Specific-UPM
Don José	Site Specific-UPM
Don Valentin	Site Specific-UPM
Dorondon	Site Specific-UPM
Eunoble	Site Specific-UPM
Grador	Site Specific-UPM
Iberus	Site Specific-UPM
Incanto	Site Specific-UPM
Kenobi	Site Specific-UPM
Kiko Nick	Site Specific-UPM
LG Confianza	Site Specific-UPM
Olivadur	Site Specific-UPM
RGT Partitur	Site Specific-UPM
RGT Rumbadur	Site Specific-UPM
SY Atlante	Site Specific-UPM
Tussur	Site Specific-UPM
GS Armando	Site Specific-UPM & HORTA
Massimo Meridio	Site Specific-UPM & HORTA
Tito Flavio	Site Specific-UPM & HORTA
Farah	Site Specific-UPM & HORTA & SIDI & Marchouch
Reale	Site Specific-UPM & SIDI & Marchouch
Burgos	Site Specific-UPM & UNIDEB
Giulio	Site Specific-UPM & UNIDEB & SIDI & Marchouch
Tirex	Site Specific-UPM & UNITUS & HORTA

Table S14: List of durum wheat varieties sown in DUS trials. 253 varieties in total, including Unibo varieties from the durum panel 1 (Maccaferri et al., 2015).

Plot	Variety Name
1	Cham-1
2	Aramon
3	Cham-9
4	Acadur
5	Kenobi
6	Cesare
7	SY_Esperto
8	Don_Norman
9	Colombo
10	Dario
11	Ottaviano
12	Provenzal
13	Altar_84
14	Core
15	Yavaros
16	Platone
17	Amira
18	Diogene
19	Kiko_Nick
20	Margherita
21	Duramonte
22	Calero
23	Incanto
24	Carioca
25	Auradur
26	San Carlo
27	Riyad
28	Fuego
29	Dylan
30	Verace
31	SY Leonardo
32	Pigreco
33	Amilcar
34	Credit
35	Tomouh
36	Emilio_Lepido
37	Bacardi
38	Caboto
39	Iride
40	Furio_Camillo
41	Avispa
42	Isildur
43	Saragolla
44	SY_Lido

45	Cordour
46	Oliver
47	Nonno_Mariano
48	Farah
49	CONTROL
50	Yasmin
51	Marakas
52	Anouar
53	Cartesio
54	Sorrento
55	Ermes
56	Creso
57	Durofinus
58	Nobilis
59	Pescadou
60	CONTROL
61	MV_Pennedur
62	Quadrato
63	Augusto
64	Meridiano
65	Tiziana
66	Prospero
67	Valforte
68	GK_Selyemdur
69	SY_Cisco
70	CONTROL
71	Secolo
72	Ramirez
73	Grazia
74	Ariosto
75	Tito Flavio
76	Linea 1
77	Plaza
78	Redidenari
79	Latino
80	Marouane
81	Eunoble
82	RGT Rangodur
83	LG Anubis
84	Minosse
85	SY Atlante
86	RGT Natur
87	Boniduro
88	Nadif
89	Lupidur
90	Panoramix
91	Brigante

92	Don Valentin
93	RGT Olivadur
94	Troubadur
95	Duilio
96	Reale
97	Don Ortega
98	Beltorax
99	Mameli
100	Nazareno
101	Sfinge
102	Qualidou
103	Babylone
104	Miradoux
105	Natal
106	Giulio
107	Orizzonte
108	Anvergur
109	Cham-3
110	GK Betadur
111	Massimo Meridio
112	Bering
113	Kanakis
114	Irden
115	Simeto
116	Calcas
117	Tussur
118	Trimulato
119	CONTROL
120	Haby
121	Anco Marzio
122	Durango
123	Federico II
124	GK Julidur
125	Ovidio
126	Duroflavus
127	Varano
128	Cincinnato
129	Zetae
130	Duromax
131	Achille
132	RGT Leondur
133	RD17236
134	Tarek
135	Levante
136	Don Pedro
137	Marco Aurelio
138	Iberus

139	Bolo
140	Arcangelo
141	Obelix
142	Cuspide
143	Opera
144	Casteldoux
145	Elsadur
146	Catasta
147	RGT Partitur
148	Diadur
149	Orobel
150	Liberdur
151	Carpio
152	Maestà
153	Casanova
154	Ginseng
155	Faraj
156	Solstizio
157	Solea
158	RGT Monbecur
159	RGT Voltadur
160	Ciccio
161	Asterix
162	Odisseo
163	Euroduro
164	Athoris
165	Dorondon
166	Brancaleone
167	Adone
168	SY Alfiere
169	Russello
170	SY Nilo
171	Diamante
172	Ourgh
173	Latinur
174	Kronos
175	Neruda
176	Don Ricardo
177	Aureo
178	Mimmo
179	Cham-5
180	Gallareta
181	Mario
182	Massara
183	Antalis
184	Matusalem
185	Don Matteo

186	MV Pelsodur
187	Edmore
188	RGT Rumbadur
189	Telemaco
190	Heraklion
191	Sculptur
192	Normanno
193	Jabato
194	Zagharin
195	Messapia
196	Waha
197	Marzak
198	LG Confianza
199	Primadur
200	Produra
201	CONTROL
202	Tirex
203	Egeo
204	Karur
205	CONTROL
206	Semperdur
207	Alemanno
208	Gibraltar
209	CONTROL
210	RGT_Aventadur
211	Neodur
212	Taranto
213	Teodorico
214	Khiaer
215	RGT_Voilur
216	Gregorio
217	RD16214
218	Karim
219	Lloyd
220	Monastir
221	Claudio
222	Murano
223	Svevo
224	Daurur
225	PR22D89
226	Cham-7
227	RGT Beticur
228	Don Josè
229	Pedroso
230	Magellano
231	Yazi_10
232	Relief

233	Nassira
234	Grador
235	Burgos
236	Reglise
237	Atoudur
238	Aneto
239	Duroi
240	Amjad
241	Martinur
242	CONTROL
243	Baronio
244	CONTROL
245	Montale
246	Mohawk
247	Amina
248	Idefix
249	Linea 2
250	Colosseo
251	Guadalso
252	GS_Armando
253	Duprì

Table S15:: Anova tables for DUS traits, reporting the significance of the phenotypic data in relation to replicates and position of each variety in the field (rows and columns).

Trait	Vraiables	Sum Sq	Df	F value	Pr(>F)	
Plant growth habit	**Genotype**	158.43	232	3.01	1.24E-14	***
Plant growth habit	**Replicate**	0.09	1	0.41	0.52	
Plant growth habit	**row**	5.61	21	1.18	0.27	
Plant growth habit	**column**	6.05	22	1.21	0.24	
Plant growth habit	**Residuals**	42.84	189			
Heading date	**Genotype**	4601.7	231	29.49	2.20E-16	***
Heading date	**Replicate**	1.3	1	1.99	0.16	
Heading date	**row**	53.9	21	3.8	3.99E-07	***
Heading date	**column**	24	22	1.61	0.047	*
Heading date	**Residuals**	126.3	187			
Culm_glaucosity_of_neck	**Genotype**	125.1	228	2.8724	5.83E-13	***
Culm_glaucosity_of_neck	**Replicate**	0.448	1	2.3474	0.1273	
Culm_glaucosity_of_neck	**row**	4.124	21	1.028	0.4324	
Culm_glaucosity_of_neck	**column**	17.256	22	4.106	5.27E-08	***
Culm_glaucosity_of_neck	**Residuals**	33.62	176			
Ear glaucosity	**Genotype**	595.8	235	6.92	2.20E-16	***
Ear glaucosity	**Replicate**	0.29	1	0.78	0.38	
Ear glaucosity	**row**	5.81	21	0.76	0.77	
Ear glaucosity	**column**	24.45	22	3.03	0.000021	***
Ear glaucosity	**Residuals**	69.97	191			
Awn color	**Genotype**	309.02	236	127.98	<2e-16	***
Awn color	**Replicate**	0.002	1	0.17	0.68	
Awn color	**row**	0.24	21	1.1	0.35	
Awn color	**column**	0.23	22	1.02	0.45	
Awn color	**Residuals**	1.98	193			
Ear colour	**Genotype**	118.47	236	117.29	<2e-16	***
Ear colour	**Replicate**	0	1	0	0.99	
Ear colour	**row**	0.09	21	0.94	0.54	
Ear colour	**column**	0.08	22	0.8	0.73	
Ear colour	**Residuals**	0.83	193			
Ear_density	**Genotype**	249.96	235	45.2	<2e-16	***
Ear_density	**Replicate**	0.02	1	0.88	0.35	
Ear_density	**row**	0.33	21	0.67	0.86	
Ear_density	**column**	0.67	22	1.3	0.18	
Ear_density	**Residuals**	4.52	192			

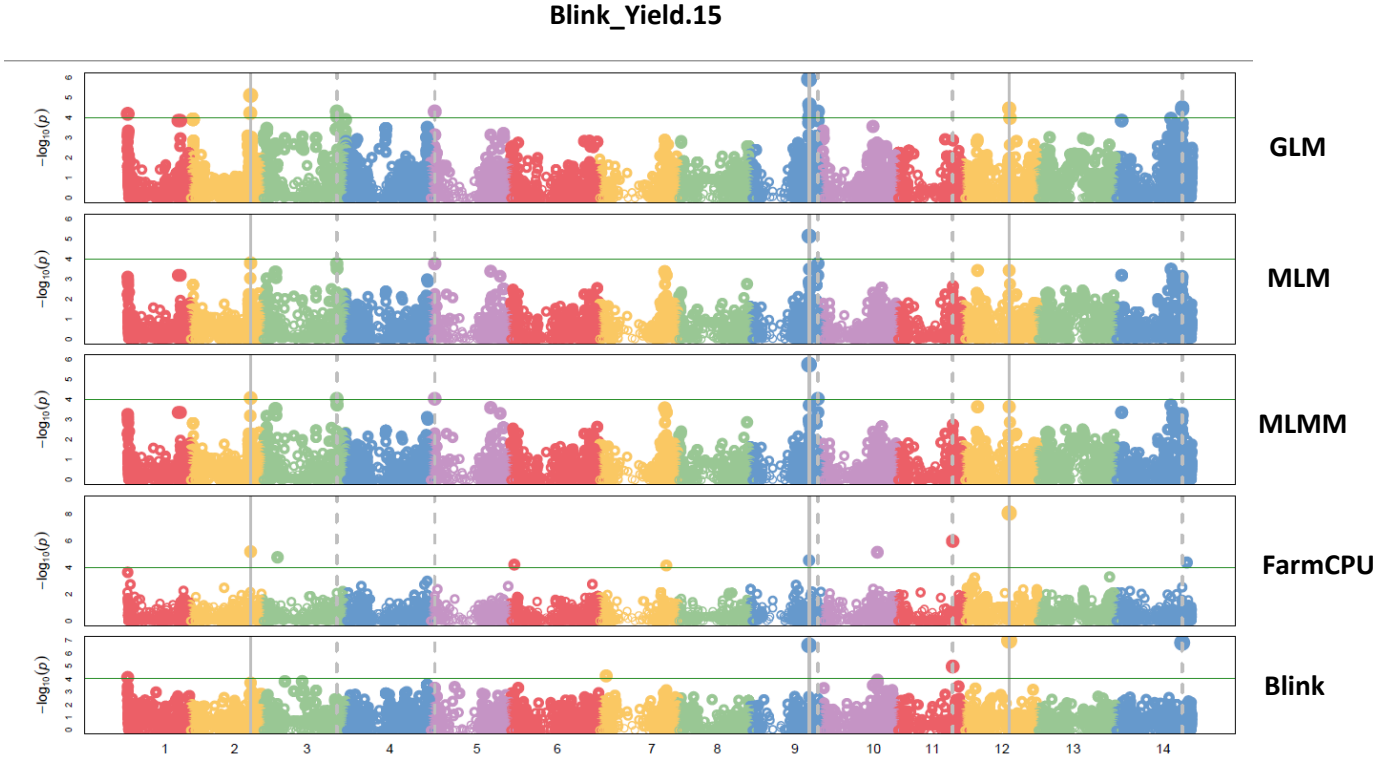


Figure S32: Total GWAS on Yield15 trait for VCU durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK

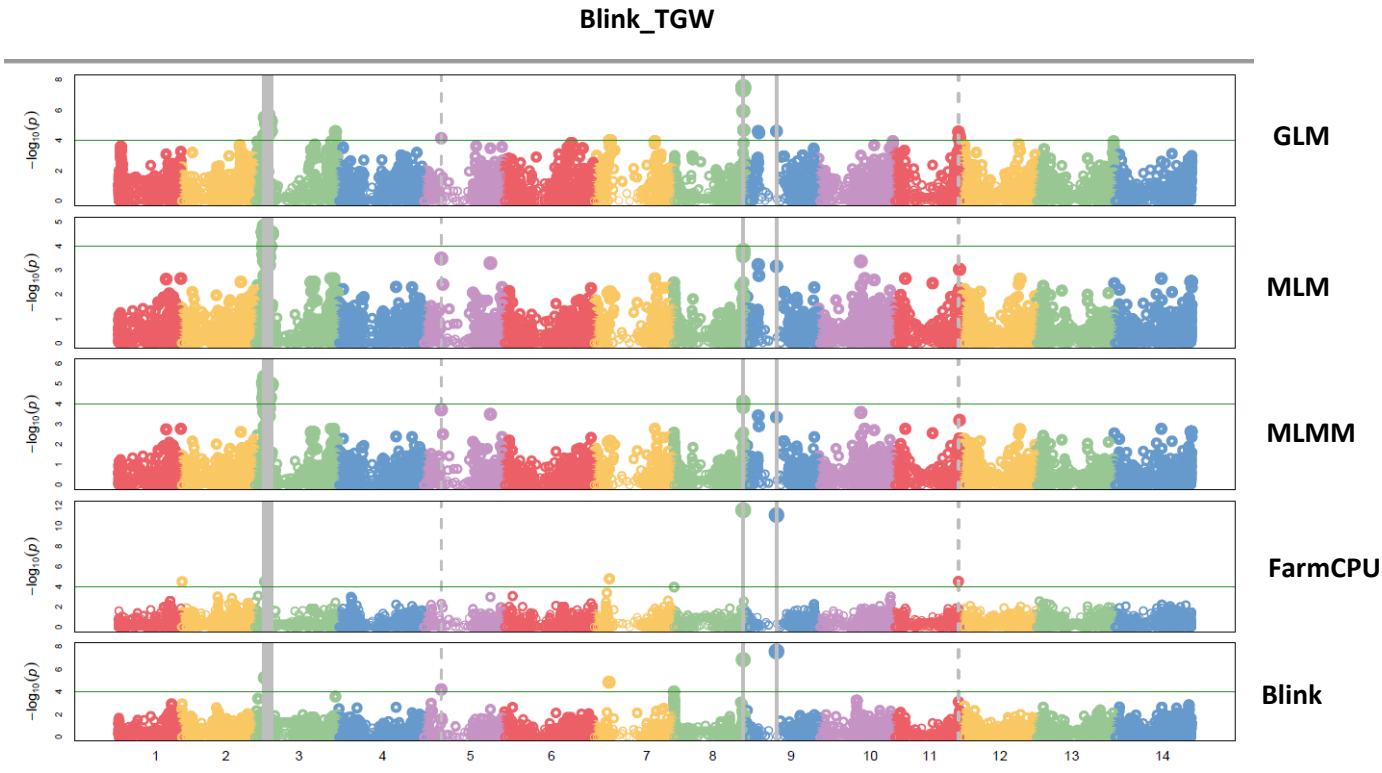


Figure S33: Total GWAS on TGW trait for VCU durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK

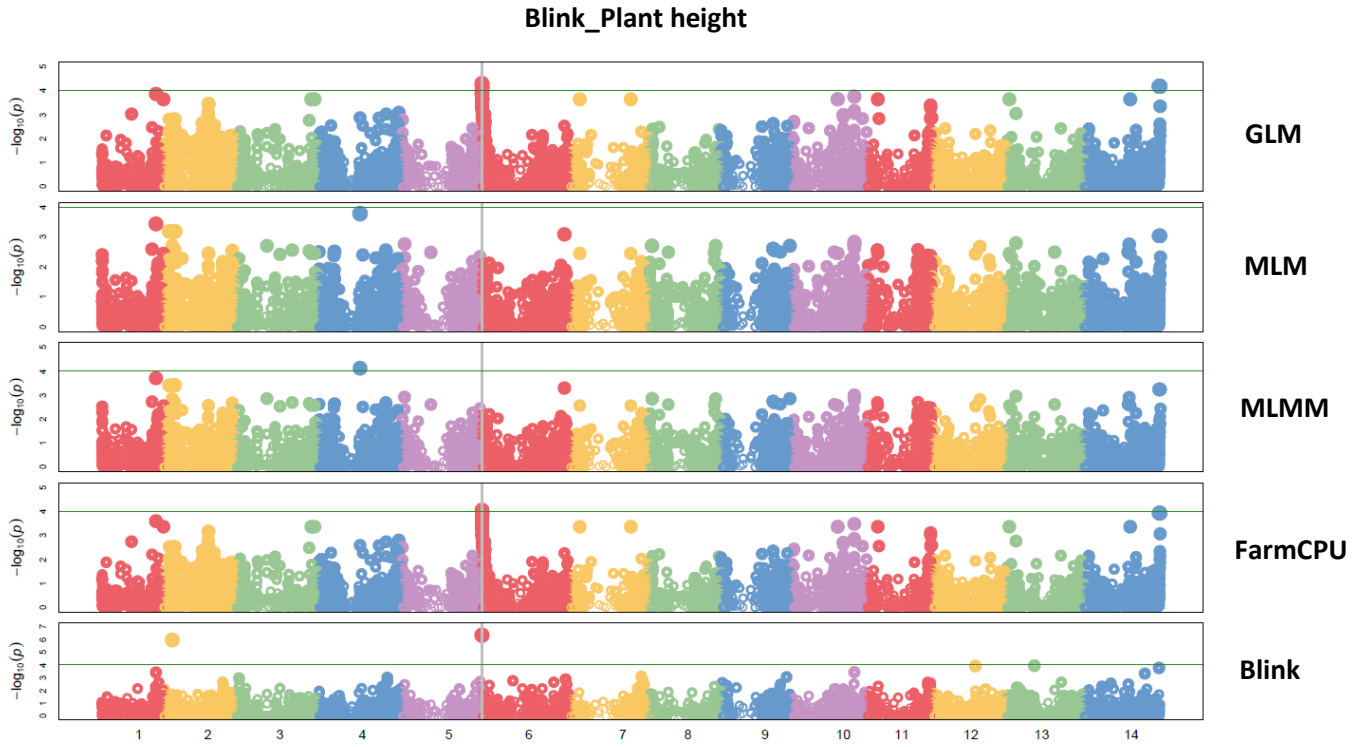


Figure S34: Total GWAS on plant height trait for VCU durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK

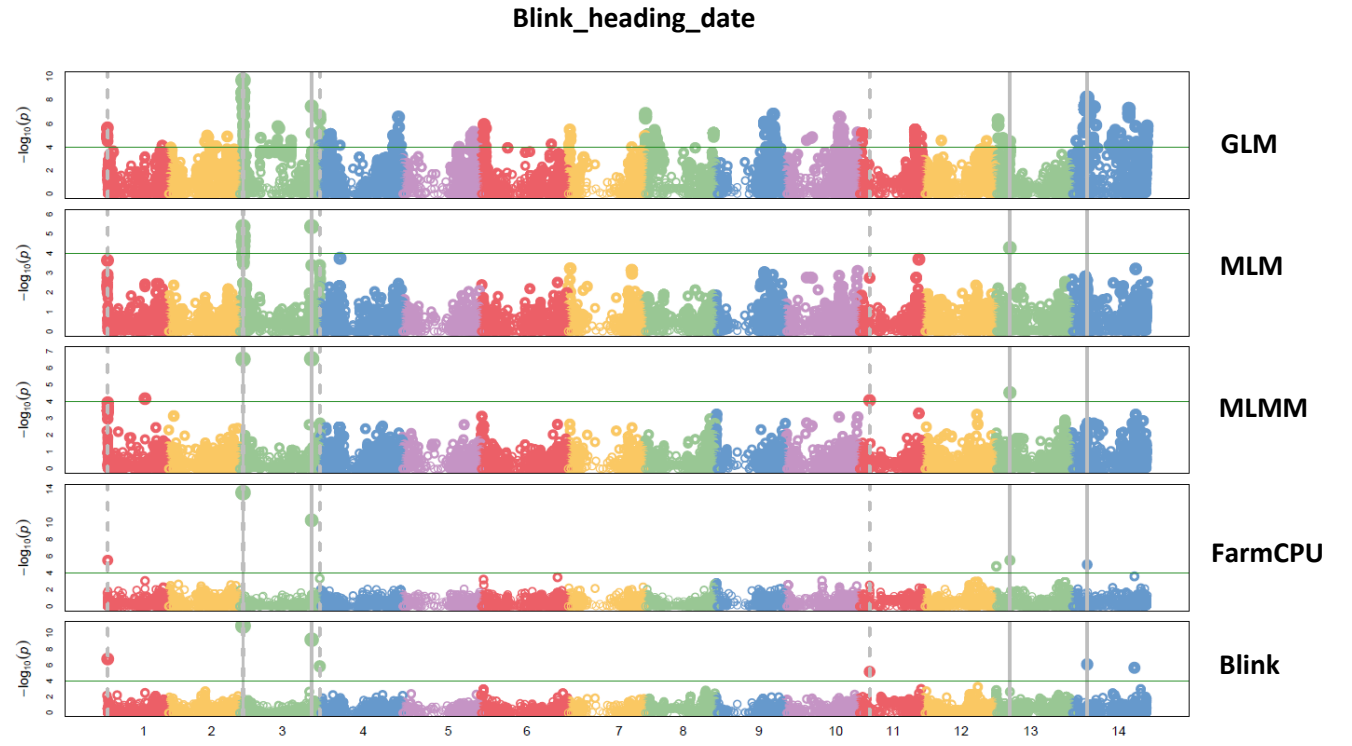


Figure S35: Total GWAS on heading date trait for VCU durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK

Table S16: Genes included in the confidence interval of QTLs on chromosomes 6B, 7B, 5A and 2A for Yield_15 detected with Blink model. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD6Bv1G132180	430836471	430840128	Receptor-kinase, putative
TRITD6Bv1G132230	431302867	431304156	Leucine-rich repeat protein kinase family protein
TRITD6Bv1G132240	431309635	431310618	
TRITD6Bv1G132250	431314486	431323464	methyl-coenzyme M reductase II subunit gamma, putative (DUF3741)
TRITD6Bv1G132290	431397122	431400784	Protein HIR1 G
TRITD6Bv1G132370	431793986	431817246	CLIP-associating family protein
TRITD6Bv1G132420	431951070	431989122	Prolyl oligopeptidase family protein
TRITD6Bv1G132500	432279851	432291668	Oligopeptide transporter, putative
TRITD6Bv1G132630	432585600	432587804	GDSL esterase/lipase G
TRITD6Bv1G132650	432589608	432602698	Polyribonucleotide nucleotidyltransferase
TRITD7Bv1G198610	622812851	622818763	BTB/POZ domain containing protein, expressed
TRITD7Bv1G198690	623239178	623239963	BTB/POZ domain containing protein
TRITD7Bv1G198720	623313659	623314069	BTB/POZ domain containing protein
TRITD7Bv1G198780	623462119	623485520	Inorganic pyrophosphatase 2
TRITD7Bv1G198820	623573035	623575356	F-box/RNI superfamily protein
TRITD7Bv1G198930	623891609	623895411	BTB/POZ domain containing protein, expressed
TRITD7Bv1G198940	623944498	623944872	Growth/differentiation factor 11 G
TRITD7Bv1G198950	623945604	623947164	RING/U-box superfamily protein G
TRITD7Bv1G198990	624077247	624078047	Germin-like protein
TRITD7Bv1G199000	624082149	624083296	co-factor for nitrate, reductase and xanthine dehydrogenase 5 G
TRITD7Bv1G199010	624083714	624084150	
TRITD7Bv1G199020	624087852	624088649	DUF674 family protein
TRITD7Bv1G199050	624126048	624126335	B3 domain-containing protein family
TRITD7Bv1G199100	624309018	624311675	Phosphoglycerate kinase
TRITD7Bv1G199120	624373584	624375972	carboxyl-terminal peptidase (DUF239)
TRITD7Bv1G199170	624492621	624493505	BTB/POZ domain containing protein, expressed
TRITD7Bv1G199230	624572559	624573865	Peroxidase
TRITD7Bv1G199260	624608291	624608581	BTB-POZ and MATH domain protein G

TRITD7Bv1G199300	624680624	624681061	BTB/POZ domain containing protein, expressed
TRITD7Bv1G199460	625154832	625156375	Peroxidase
TRITD7Bv1G199470	625164187	625166161	Zinc-binding protein
TRITD7Bv1G199540	625269881	625271244	GDSL esterase/lipase
TRITD7Bv1G199550	625336602	625337088	NADH-ubiquinone oxidoreductase chain 1
TRITD7Bv1G199580	625343391	625437320	Kinase-like protein
TRITD5Av1G203320	545841174	545854435	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein, putative
TRITD5Av1G203330	545855331	545856238	Photosystem II protein
TRITD5Av1G203340	545860098	545869249	Myosin family protein, putative, expressed
TRITD5Av1G203460	546226501	546230147	Evolutionarily conserved C-terminal region 2
TRITD5Av1G203470	546237669	546238490	Short-chain dehydrogenase/reductase family protein
TRITD5Av1G203500	546260618	546263962	PHD finger protein ING
TRITD5Av1G203520	546265571	546266881	Protein NDH-DEPENDENT CYCLIC ELECTRON FLOW 5
TRITD5Av1G203530	546267757	546268722	Receptor-like protein kinase
TRITD5Av1G203580	546323261	546324726	Receptor-like kinase
TRITD5Av1G203590	546328844	546332006	Protein kinase family protein, putative, expressed
TRITD5Av1G203600	546332763	546333660	NAD(P)H dehydrogenase (Quinone)
TRITD5Av1G203610	546333971	546334601	Maltase-glucoamylase, intestinal G
TRITD5Av1G203630	546364148	546367474	B3 domain-containing protein
TRITD5Av1G203640	546429468	546433662	Cytosolic Fe-S cluster assembly factor NAR1
TRITD5Av1G203670	546488084	546494803	ATP-dependent RNA helicase, putative
TRITD5Av1G203680	546526930	546530494	RNA-binding protein 39
TRITD5Av1G203690	546531451	546535150	Glycosyltransferase
TRITD5Av1G203730	546684717	546689207	Nitrate transporter 1.1 G
TRITD5Av1G203740	546695256	546696656	Ammonium transporter
TRITD5Av1G203760	546752290	546948030	Beta-glucosidase
TRITD5Av1G203770	546801393	546801860	Beta-glucosidase
TRITD5Av1G203790	546814761	546820749	F-box family protein
TRITD5Av1G203820	546830023	546831993	Pentatricopeptide repeat-containing protein
TRITD5Av1G203830	546930411	546933841	Receptor-like protein kinase
TRITD5Av1G203860	546964919	547140019	Beta-glucosidase
TRITD5Av1G203890	547140986	547152548	Protein kinase
TRITD5Av1G203900	547153202	547154786	Protein-lysine N-methyltransferase

TRITD5Av1G203910	547155317	547165350	Universal stress protein family protein
TRITD5Av1G203920	547167673	547169236	Universal stress protein family protein
TRITD5Av1G203930	547169913	547173129	Tetratricopeptide repeat-containing family protein
TRITD5Av1G203970	547193682	547201269	Kinesin-like protein
TRITD5Av1G204010	547228234	547230893	Strictosidine synthase
TRITD5Av1G204040	547348643	547352985	Transcriptional adapter ADA2
TRITD5Av1G204050	547372613	547375588	Structural maintenance of chromosomes protein 2 G
TRITD5Av1G204120	547477973	547478648	HMG-Y-related protein A
TRITD5Av1G204190	547740839	547742549	Glucan endo-1,3-beta-glucosidase-like protein
TRITD5Av1G204260	547880121	547880582	CDO504
TRITD5Av1G204270	547881732	547883302	Glucan endo-1,3-beta-glucosidase 1
TRITD5Av1G204280	547884184	547888883	RAN GTPase-activating protein 1
TRITD5Av1G204300	547889340	547892515	Oligopeptide transporter, putative
TRITD5Av1G204320	547912562	547915197	Peroxisomal membrane protein 2 G
TRITD2Av1G043170	94092206	94092616	
TRITD2Av1G043180	94099305	94099637	Phosphate carrier protein, mitochondrial
TRITD2Av1G043240	94211740	94215254	Peroxidase
TRITD2Av1G043250	94214495	94214851	Peroxidase 1 G
TRITD2Av1G043300	94336183	94336521	Glutaredoxin, putative
TRITD2Av1G043500	94953246	94961150	Leucine--tRNA ligase
TRITD2Av1G043580	95042901	95045768	Phosphoserine phosphatase
TRITD2Av1G043590	95063116	95064607	Peroxidase
TRITD2Av1G043630	95120665	95121520	Peroxiredoxin
TRITD2Av1G043640	95125162	95127117	tolB protein-like protein
TRITD2Av1G043650	95133908	95146519	Kinesin-like protein
TRITD2Av1G043700	95215880	95221883	Serine/threonine-protein kinase haspin
TRITD2Av1G043710	95223274	95226078	SNARE

Table S17: Genes included in the confidence interval of QTLs on chromosomes 5A, 4B and 2A for TGW detected with Blink model. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD5Av1G097030	278798396	278799910	Vacuolar protein sorting-associated protein 45 like G
TRITD5Av1G097050	278888447	278888902	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase G
TRITD5Av1G097060	278890137	278891936	Pentatricopeptide repeat-containing protein
TRITD5Av1G097200	279485390	279490053	Bifunctional uridylyltransferase/uridylyl-removing enzyme
TRITD5Av1G097390	280136432	280141094	Pentatricopeptide repeat-containing protein
TRITD5Av1G097420	280145926	280147492	Alpha/beta-Hydrolases superfamily protein
TRITD5Av1G097690	281065529	281066396	Terminal flower 1-like protein
TRITD4Bv1G194640	645229944	645234946	Kinase family protein
TRITD4Bv1G194690	645315282	645325658	Endoribonuclease Dicer-like protein 3
TRITD4Bv1G194700	645326240	645327520	Nascent polypeptide-associated complex alpha subunit, putative, expressed
TRITD4Bv1G194710	645456949	645458447	Syntaxin, putative
TRITD4Bv1G194720	645463642	645464923	Peroxidase
TRITD4Bv1G194730	645468053	645471653	Peroxidase
TRITD4Bv1G194740	645475396	645479466	TVP38/TMEM64 family membrane protein
TRITD4Bv1G194790	645552186	645553073	Hfr-2-like protein
TRITD4Bv1G194800	645590804	645679455	Hfr-2-like protein
TRITD4Bv1G194830	645714211	645716077	Hfr-2-like protein
TRITD4Bv1G194860	645753770	645755561	Hfr-2-like protein
TRITD4Bv1G194870	645807324	645809610	Amino acid transporter, putative
TRITD4Bv1G194910	645868118	645869590	NBS-LRR-like resistance protein

TRITD4Bv1G194950	645907753	645908983	Aquaporin-like protein
TRITD4Bv1G195040	646044107	646044784	Protein F12F1.11-, putative G
TRITD4Bv1G195050	646051695	646053593	DUF674 family protein
TRITD4Bv1G195060	646075231	646077274	DUF674 family protein
TRITD4Bv1G195080	646079007	646082418	Remorin
TRITD4Bv1G195100	646098386	646098826	DUF538 family protein (Protein of unknown function, DUF538)
TRITD4Bv1G195150	646201091	646202401	Heavy metal transport/detoxification superfamily protein
TRITD4Bv1G195200	646335984	646340715	autoinhibited Ca2+ -ATPase, isoform 8 G
TRITD4Bv1G195210	646361933	646362463	Non-specific lipid-transfer protein
TRITD4Bv1G195230	646421171	646434994	Purple acid phosphatase
TRITD4Bv1G195320	646647334	646648293	B3 domain-containing protein
TRITD4Bv1G195390	646798562	646800419	Hfr-2-like protein
TRITD4Bv1G195430	646865352	646867163	Leucine-rich repeat receptor-like protein kinase family protein
TRITD4Bv1G195500	646915682	646917341	Receptor-kinase, putative
TRITD4Bv1G195510	646917525	646919684	Leucine-rich repeat receptor-like protein kinase family protein
TRITD4Bv1G195530	646962125	646965692	Receptor-kinase, putative
TRITD4Bv1G195580	647044051	647044362	Histone H4
TRITD4Bv1G195600	647121341	647121652	Histone H4
TRITD4Bv1G195620	647132132	647136077	NAC domain protein,
TRITD4Bv1G195630	647168920	647171460	F12P19.7, putative isoform 2 G
TRITD4Bv1G195640	647172244	647174581	ALA-interacting subunit 3
TRITD4Bv1G195750	647478767	647479834	NBS-LRR-like resistance protein
TRITD4Bv1G195760	647483582	647485339	Ubiquitin
TRITD4Bv1G195800	647577444	647577656	Eukaryotic translation initiation factor 4G, putative isoform 2 G
TRITD2Av1G043170	94092206	94092616	
TRITD2Av1G043180	94099305	94099637	Phosphate carrier protein, mitochondrial

TRITD2Av1G043240	94211740	94215254	Peroxidase
TRITD2Av1G043250	94214495	94214851	Peroxidase 1 G
TRITD2Av1G043300	94336183	94336521	Glutaredoxin, putative
TRITD2Av1G043500	94953246	94961150	Leucine--tRNA ligase
TRITD2Av1G043580	95042901	95045768	Phosphoserine phosphatase
TRITD2Av1G043590	95063116	95064607	Peroxidase
TRITD2Av1G043630	95120665	95121520	Peroxiredoxin
TRITD2Av1G043640	95125162	95127117	tolB protein-like protein
TRITD2Av1G043650	95133908	95146519	Kinesin-like protein
TRITD2Av1G043700	95215880	95221883	Serine/threonine-protein kinase haspin
TRITD2Av1G043710	95223274	95226078	SNARE

Table S18: Genes included in the confidence interval of QTLs on chromosomes 3B, and 1B for plant height detected with Blink model. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD3Bv1G000010	19320	21113	Transcription factor
TRITD3Bv1G000040	58017	60765	Ankyrin repeat protein family-like protein
TRITD3Bv1G000080	300930	303306	Ring box family protein
TRITD3Bv1G000100	304233	309067	1-deoxy-D-xylulose 5-phosphate reductoisomerase
TRITD3Bv1G000190	396335	396985	Transcription initiation factor TFIID subunit 4B G
TRITD3Bv1G000200	399980	401050	Transcription elongation factor SPT5 G
TRITD3Bv1G000220	508869	511463	Serine/threonine-protein kinase ATM G
TRITD3Bv1G000230	535358	535749	Cysteine-rich protein
TRITD3Bv1G000240	539365	540584	Rer1 protein, putative
TRITD3Bv1G000290	706642	712029	Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase
TRITD3Bv1G000300	718265	719851	Polycomb group protein EMBRYONIC FLOWER 2 G
TRITD3Bv1G000310	1090268	1094704	Chaperone protein DnaJ
TRITD3Bv1G000320	1096291	1099576	Subtilisin-like protease
TRITD3Bv1G000330	1101875	1103297	Isoflavone reductase-like protein

TRITD3Bv1G000350	1108938	1110039	Subtilisin-like protease
TRITD3Bv1G000360	1114910	1118298	Subtilisin-like protease
TRITD3Bv1G000370	1125787	1128864	Subtilisin-like protease
TRITD3Bv1G000420	1186382	1187310	Isoflavone reductase-like protein
TRITD3Bv1G000450	1207355	1237623	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000460	1229055	1231026	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000470	1251849	1254384	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000480	1271090	1275023	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000510	1323477	1326192	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000530	1381936	1384059	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000550	1401612	1409953	Myosin-1 G
TRITD3Bv1G000560	1418574	1421315	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000570	1430052	1432634	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000580	1444435	1446598	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000590	1456688	1457802	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000610	1469038	1470723	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD3Bv1G000620	1479129	1480397	F-box like protein
TRITD1Bv1G030710	81760426	81761583	AT5G11810-like protein G
TRITD1Bv1G030900	82473590	82475847	WRKY transcription factor
TRITD1Bv1G030970	82703303	82705602	Aminotransferase-related family protein
TRITD1Bv1G030980	82721836	82723437	Kinase family protein

Table S19: Genes included in the confidence interval of QTLs on chromosomes 2A, and 1A for heading date detected with Blink model. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD2Av1G019050	36289642	36293816	Kinase interacting (KIP1-like) family protein
TRITD2Av1G019060	36313105	36314157	Expansin protein
TRITD2Av1G019080	36327882	36328444	Ribosome production factor 2-like protein G
TRITD2Av1G019100	36350313	36351698	UDP-glycosyltransferase
TRITD2Av1G019110	36356017	36359578	Expansin protein
TRITD2Av1G019120	36363569	36364523	Expansin protein
TRITD2Av1G019130	36369506	36381904	Low affinity potassium transport system protein kup G
TRITD2Av1G019170	36444868	36446845	Transcription factor protein
TRITD2Av1G019180	36452268	36455012	Protein kinase-like protein
TRITD2Av1G019200	36522474	36523052	RING/U-box superfamily protein
TRITD2Av1G019210	36526161	36535831	GRF zinc finger family protein, expressed TE?
TRITD2Av1G019220	36537714	36543948	E3 ubiquitin-protein ligase SINA-like 10 G
TRITD2Av1G019230	36558965	36559923	40S ribosomal protein S7 G
TRITD2Av1G019240	36562771	36563358	
TRITD2Av1G019250	36567124	36570283	Pseudo-response regulator
TRITD2Av1G019290	36691780	36696573	Hippocampus abundant transcript-like protein 1 G
TRITD2Av1G019320	36744246	36750772	Sporulation protein RMD1
TRITD2Av1G019330	36751677	36755494	Ascorbate peroxidase
TRITD2Av1G019340	36756356	36758653	Arogenate dehydratase
TRITD2Av1G019400	36873797	36874456	Tapetum determinant 1 G
TRITD2Av1G019430	36890646	36892887	UvrABC system protein C
TRITD2Av1G019440	36926020	36927249	Serine/threonine-protein kinase ATM G
TRITD2Av1G019460	36959198	36960913	Glycosyltransferases
TRITD2Av1G019470	37074527	37076234	F-box family protein
TRITD2Av1G019520	37293639	37384029	Elongation factor 1-alpha
TRITD2Av1G019530	37376072	37376969	Powder tolerance-related protein G
TRITD2Av1G019540	37386189	37387261	Yellow stripe-like transporter 12
TRITD2Av1G019550	37386220	37386411	Carbon storage regulator homolog G
TRITD2Av1G019570	37397770	37408859	MLO-like protein
TRITD2Av1G019650	37586613	37587843	Peroxidase
TRITD2Av1G019670	37606511	37611903	Phosphoinositide phospholipase C
TRITD2Av1G019740	37640461	37641691	Peroxidase
TRITD2Av1G019750	37663248	37673090	Symplekin

TRITD2Av1G019760	37684713	37686053	Eukaryotic aspartyl protease family protein
TRITD2Av1G019840	37830355	37831059	D-ribose-binding periplasmic
TRITD2Av1G019850	37885597	37953239	D-ribose-binding periplasmic
TRITD2Av1G019860	37981950	37982675	D-ribose-binding periplasmic
TRITD2Av1G019870	37986015	37988502	Omega-3 fatty acid desaturase
TRITD2Av1G019880	38038172	38041654	Chaperone protein dnaJ
TRITD2Av1G020040	38302371	38303667	RING/FYVE/PHD zinc finger protein
TRITD2Av1G020050	38329637	38339494	Syntaxin-binding protein 5-like protein
TRITD2Av1G020090	38399235	38399555	
TRITD2Av1G020100	38401754	38403517	disease resistance family protein / LRR family protein
TRITD2Av1G020110	38403825	38410194	NBS-LRR disease resistance protein
TRITD1Av1G003550	7703344	7705075	F-box protein
TRITD1Av1G003590	7756407	7761203	C2 calcium/lipid-binding and GRAM domain protein
TRITD1Av1G003700	8024424	8025167	Chaperone protein dnaJ G
TRITD1Av1G003710	8025610	8030231	Tubulin-specific chaperone cofactor E-like protein
TRITD1Av1G003720	8032928	8036877	Ras-like protein
TRITD1Av1G003750	8060307	8060825	Ras family protein
TRITD1Av1G003760	8088128	8089433	12-oxophytodienoate reductase-like protein
TRITD1Av1G003770	8092160	8093840	12-oxophytodienoate reductase-like protein
TRITD1Av1G003780	8124364	8126102	12-oxophytodienoate reductase-like protein
TRITD1Av1G003790	8129395	8129778	12-oxophytodienoate reductase
TRITD1Av1G003820	8139221	8147840	Serine/threonine-protein kinase
TRITD1Av1G003850	8153993	8156671	Serine/threonine-protein kinase
TRITD1Av1G003880	8191724	8192234	Serine/threonine protein phosphatase 7 long form isogeny
TRITD1Av1G003930	8461903	8481243	Leucine-rich repeat (LRR) family protein
TRITD1Av1G003970	8481286	8482743	NBS-LRR disease resistance protein-like protein
TRITD1Av1G003980	8510588	8515524	wall-associated receptor kinase-like protein
TRITD1Av1G004000	8562438	8565188	NBS-LRR disease resistance protein, putative, expressed
TRITD1Av1G004010	8624885	8625097	Keratin, type I cytoskeletal 28 G
TRITD1Av1G004030	8814754	8817837	Protein kinase family protein
TRITD1Av1G004040	8821345	8822143	Glutathione S-transferase

TRITD1Av1G004050	8823330	8828533	Phosphatidate cytidylyltransferase
TRITD1Av1G004060	8843881	8885081	RNA binding protein
TRITD1Av1G004070	8851084	8851554	Protein kinase family protein
TRITD1Av1G004080	8852105	8853681	Serine/threonine-protein kinase
TRITD1Av1G004090	8966607	8967707	Glutathione S-transferase
TRITD1Av1G004100	9067542	9069430	transmembrane protein, putative (DUF247)
TRITD1Av1G004110	9075137	9079045	Receptor-like protein kinase
TRITD1Av1G004190	9200751	9213751	NBS-LRR disease resistance protein, putative
TRITD1Av1G004210	9217421	9222451	Receptor-like kinase
TRITD1Av1G004220	9222484	9225714	Receptor-like kinase protein
TRITD1Av1G004250	9244432	9246921	transmembrane protein, putative (DUF594)
TRITD1Av1G004260	9251626	9258710	Zinc finger protein
TRITD1Av1G004300	9367138	9368722	Nuclear inhibitor of protein phosphatase 1
TRITD1Av1G004310	9371347	9375442	Disease resistance protein RPM1
TRITD1Av1G004320	9428539	9431467	Protein phosphatase 2C family protein
TRITD1Av1G004380	9719981	9721468	Flavonoid 3-O-glycosyltransferase
TRITD1Av1G004400	9749932	9751548	UDP-glycosyltransferase
TRITD1Av1G004410	9817406	9820293	WAT1-related protein
TRITD1Av1G004470	9873750	9876781	
TRITD1Av1G004480	9891355	9893751	WAT1-related protein
TRITD1Av1G004500	9923555	9923899	Cytochrome P450 protein
TRITD1Av1G004560	9960116	9964051	Disease resistance protein RPM1
TRITD1Av1G004570	10104725	10109492	Disease resistance protein RPM1
TRITD1Av1G004600	10114739	10125010	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain
TRITD1Av1G004620	10134112	10135963	Actin
TRITD1Av1G004630	10136804	10138991	Chaperone protein DnaJ
TRITD1Av1G004640	10152405	10159033	Phosphoinositide phosphatase family protein
TRITD1Av1G004650	10203150	10204852	Cytochrome P450 family protein, expressed
TRITD1Av1G004660	10217747	10218190	Bowman-Birk type trypsin inhibitor G
TRITD1Av1G004680	10234306	10234690	Trypsin inhibitor
TRITD1Av1G004700	10251229	10252608	Anthocyanin 5-aromatic acyltransferase
TRITD1Av1G004710	10256066	10260618	Mechanosensitive ion channel family protein

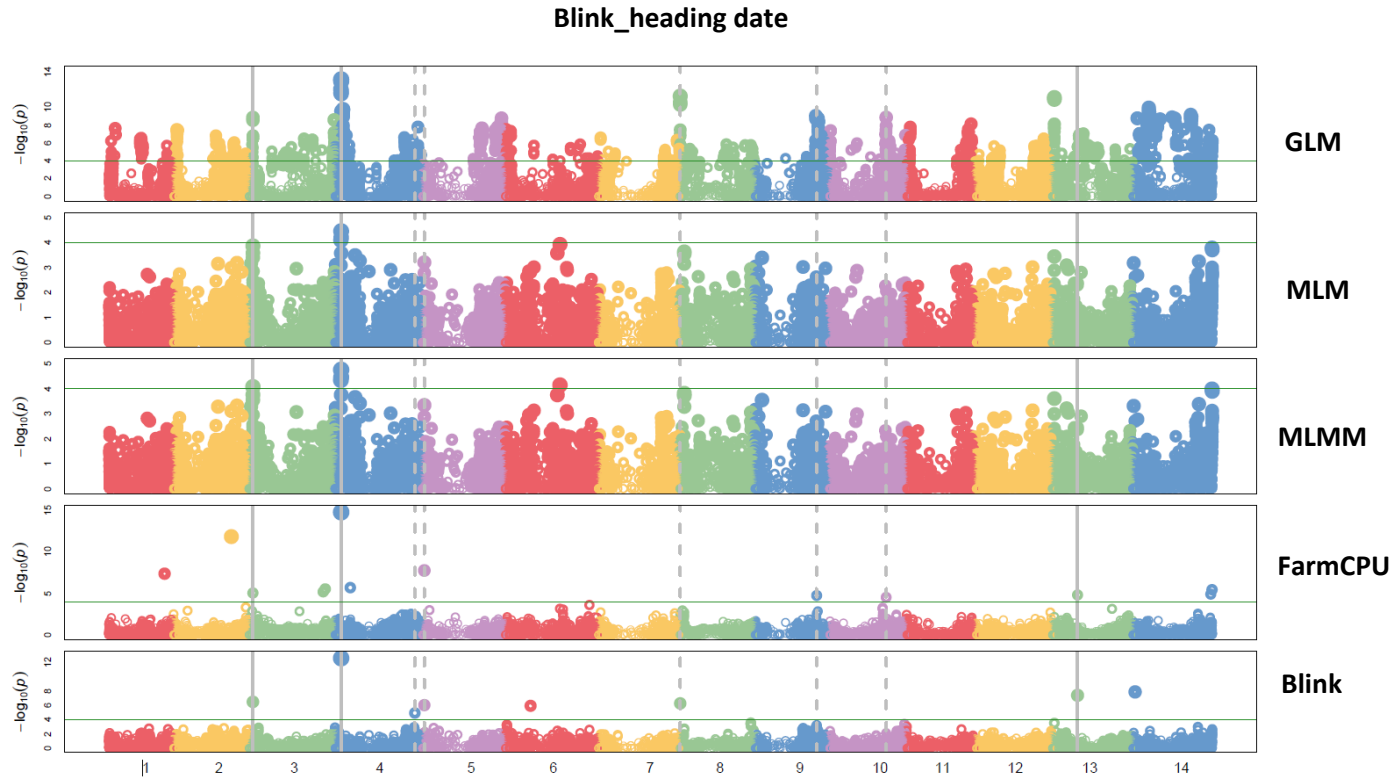


Figure S36: Total GWAS on heading date trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK

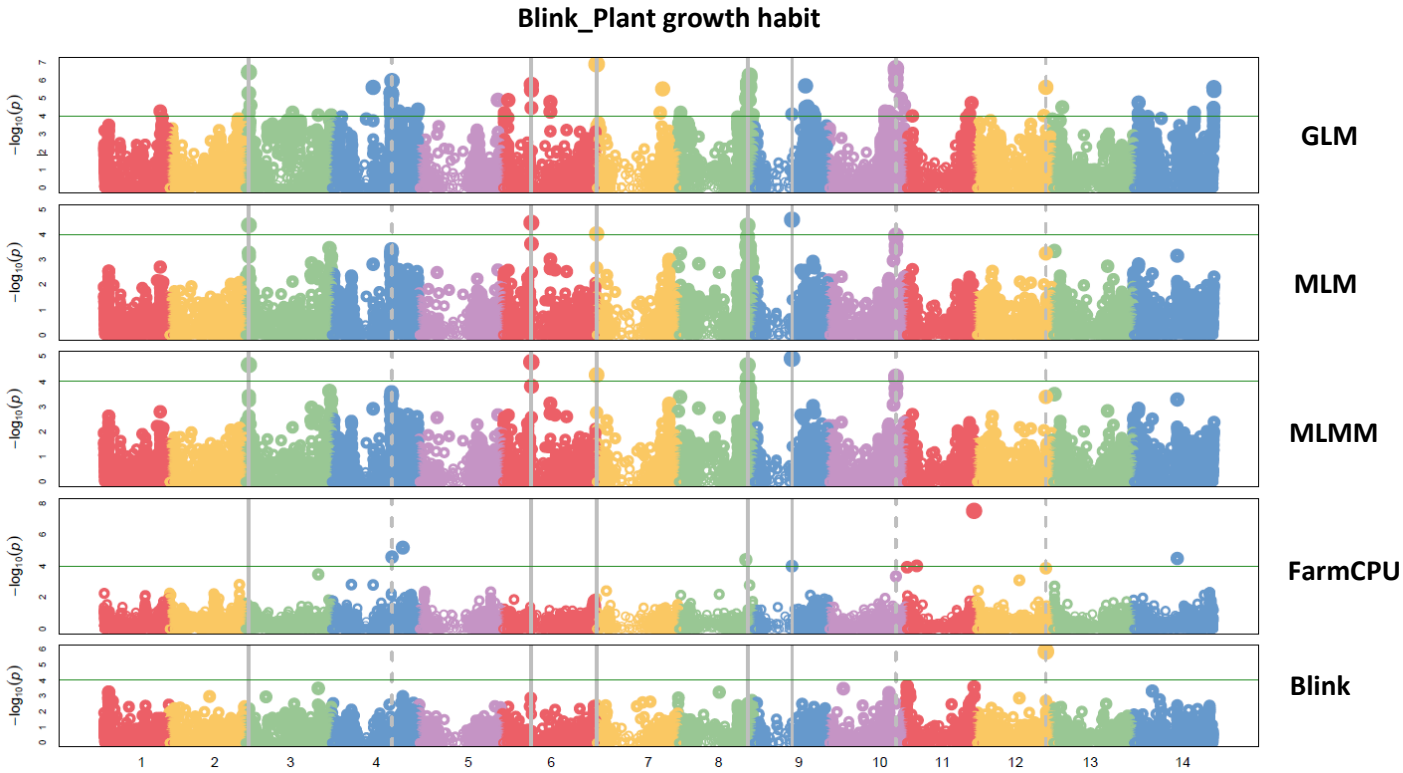


Figure S37: Total GWAS on plant growth habit trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK.

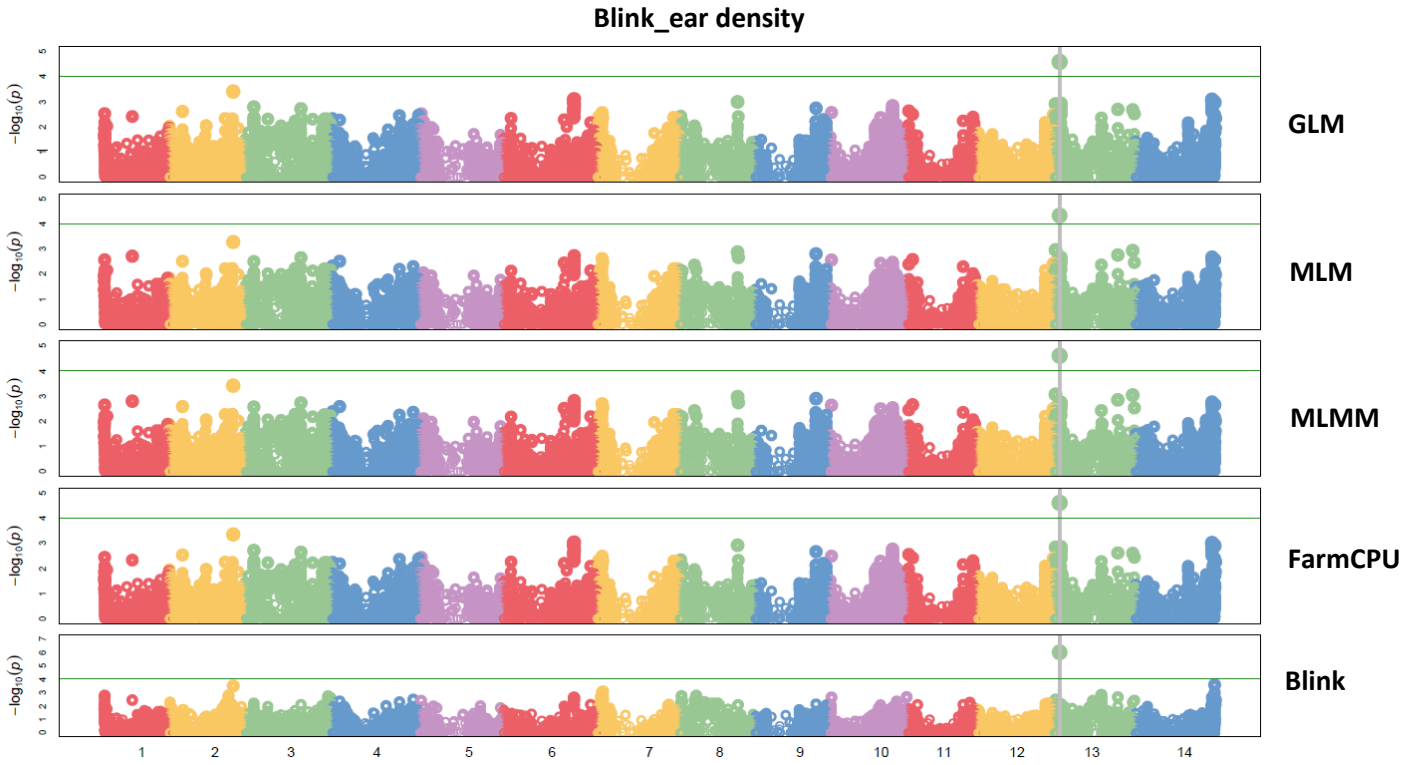


Figure S38: Total GWAS on ear density trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK.

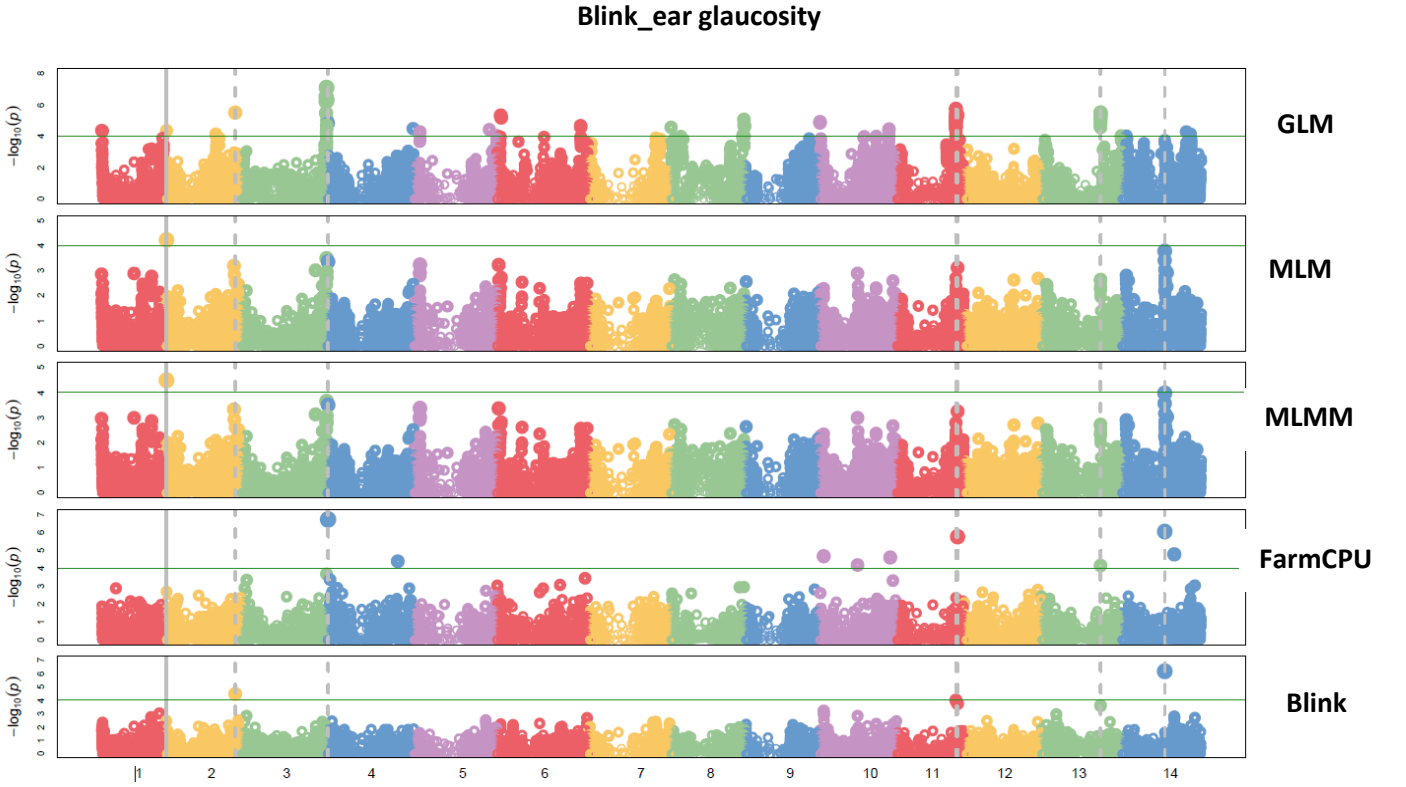


Figure S39: Total GWAS on ear glaucosity trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK.

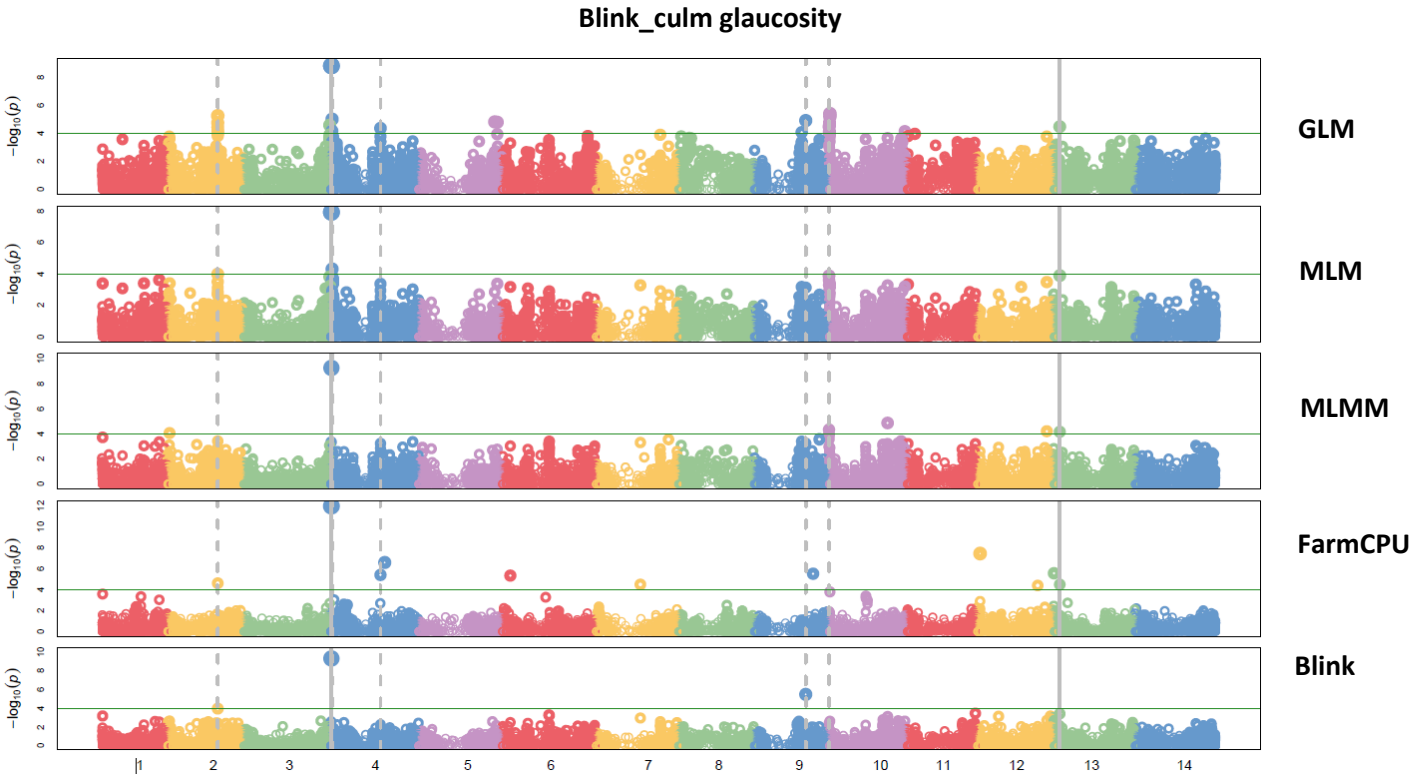


Figure S40: Total GWAS on culm glaucosity of the neck trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK.

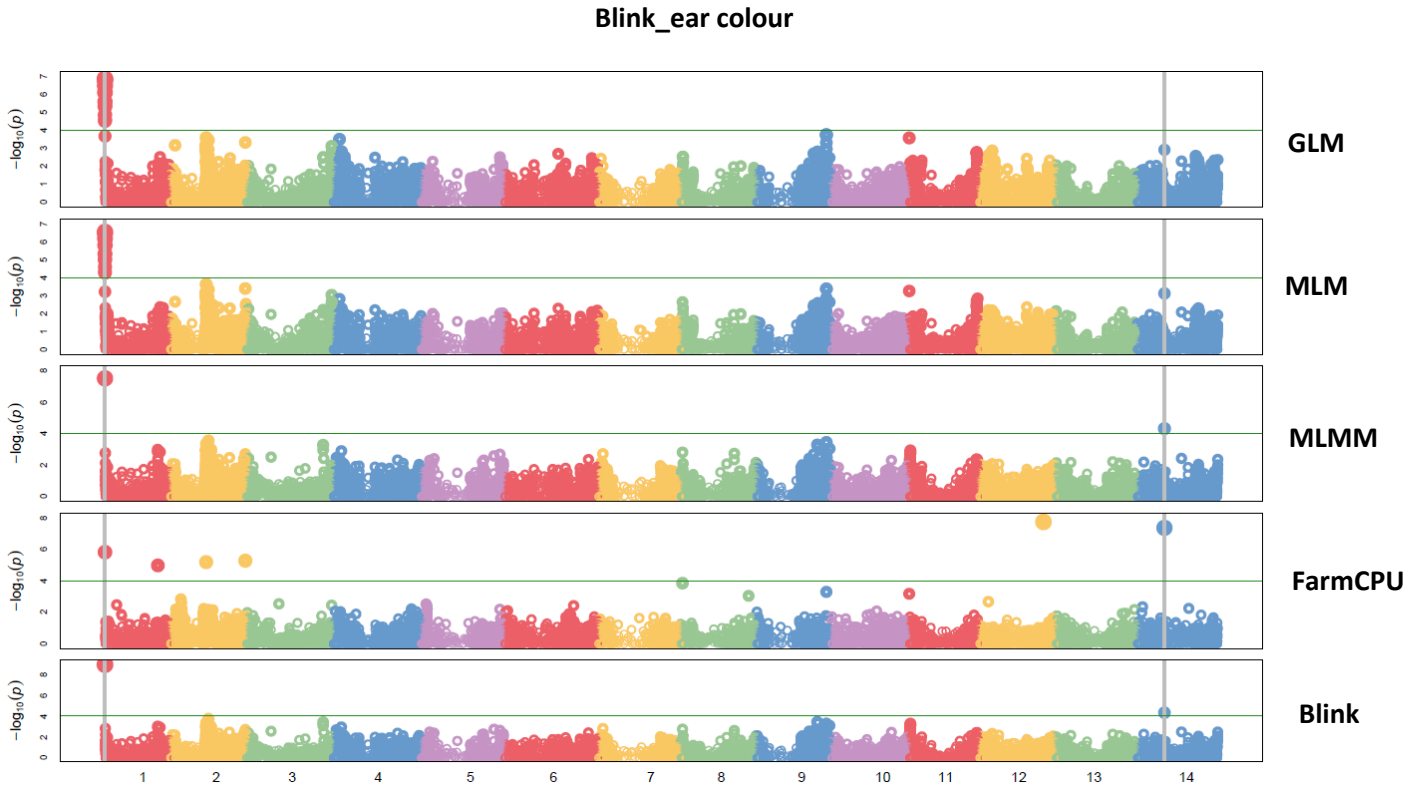


Figure S41: Total GWAS on ear colour trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK.

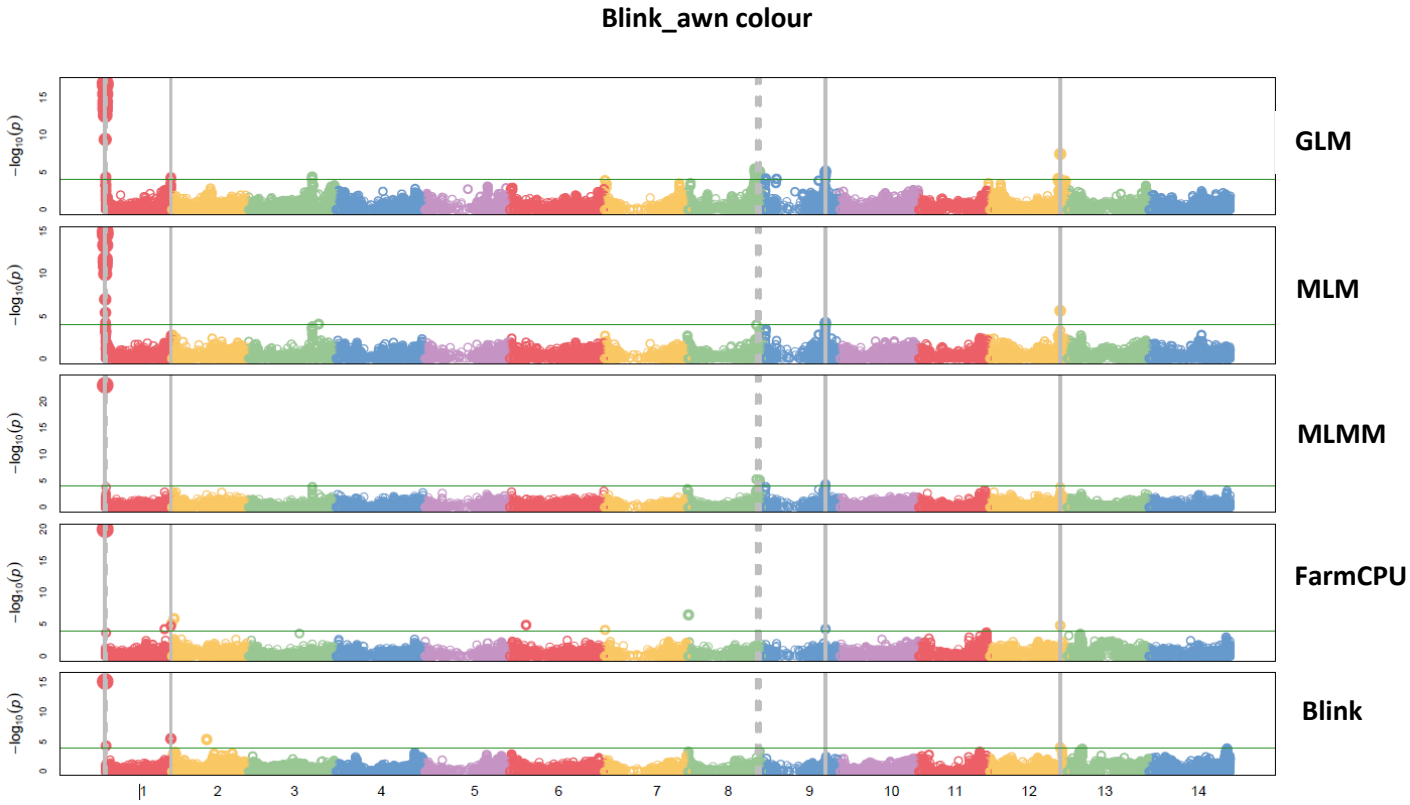


Figure S42: Total GWAS on awn colour trait for DUS durum panel. All the tested GAPIT models by GWAS + K are reported, namely (from the top to the bottom): GLM, MLM, MLMM, FarmCPU and BLINK.

Table S20: Genes included in the confidence interval of QTLs on chromosomes 2B, 7B, 7A and 2A heading date detected with Blink model in DUS dataset. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD2Bv1G025260	55902135	55906489	Kinase interacting (KIP1-like) family protein
TRITD2Bv1G025290	55922291	55922968	Expansin protein
TRITD2Bv1G025300	55925560	55926680	Expansin protein
TRITD2Bv1G025310	55929180	55931235	Expansin protein
TRITD2Bv1G025330	55932837	55934216	UDP-Glycosyltransferase
TRITD2Bv1G025340	56011401	56013178	Transcription factor

TRITD2Bv1G025350	56053235	56055910	Protein kinase-like protein
TRITD2Bv1G025480	56530134	56533829	Peroxidase family protein
TRITD2Bv1G025490	56534421	56536712	Arogenate dehydratase
TRITD2Bv1G025510	56566914	56567581	Tapetum determinant 1 G
TRITD2Bv1G025530	56650689	56652921	UvrABC system protein C
TRITD2Bv1G025540	**56658223**	**56660088**	**Serine/threonine-protein kinase ATM**
TRITD2Bv1G025590	56772170	56774159	Glycosyltransferases
TRITD2Bv1G025600	56779627	56781372	Glycosyltransferases
TRITD2Bv1G025650	56953213	56954836	F-box family protein
TRITD2Bv1G025720	57177759	57258373	Elongation factor 1-alpha
TRITD2Bv1G025740	57262710	57266846	Yellow stripe-like transporter 12
TRITD2Bv1G025750	57262741	57262932	Urease subunit gamma G
TRITD2Bv1G025770	57270268	57273322	Yellow stripe-like transporter 12
TRITD2Bv1G025780	57275377	57390728	MLO-like protein
TRITD2Bv1G025970	57588408	57589539	Peroxidase
TRITD2Bv1G025980	57599207	57602656	Phosphoinositide phospholipase C
TRITD2Bv1G025990	57613668	57614117	Peroxidase
TRITD2Bv1G026090	57682244	57691858	Symplekin
TRITD7Bv1G008090	21153198	21153935	F-box family protein

TRITD7Bv1G008120	21268182	21271622	Peroxyureidoacrylate/ureidoacrylate amidohydrolase RutB
TRITD7Bv1G008260	21614319	21621748	Serine protease HTRA1
TRITD7Bv1G008280	21708082	21709110	Eukaryotic translation initiation factor 3 subunit L G
TRITD7Bv1G008480	22075690	22080665	Alpha/beta-Hydrolases superfamily protein
TRITD7Bv1G008490	**22089866**	**22091774**	**Late embryogenesis abundant protein**
TRITD7Bv1G008520	**22093924**	**22094451**	**F-box-like protein**
TRITD7Bv1G008530	22143567	22144733	F-box protein
TRITD7Bv1G008550	22248507	22251101	F-box protein
TRITD7Bv1G008640	22478881	22483084	Protein disulfide-isomerase
TRITD7Bv1G008650	22484334	22497915	Subtilisin-like protease 1
TRITD7Bv1G008690	22681589	22682828	Peroxidase
TRITD7Bv1G008750	22856531	22857782	Peroxidase
TRITD7Bv1G008760	22883092	22885218	Galactoside 2-alpha-L-fucosyltransferase
TRITD7Bv1G008890	23094324	23095407	Dirigent protein
TRITD7Bv1G008900	23137896	23138985	Dirigent protein
TRITD7Bv1G008910	23203063	23324421	Dirigent protein
TRITD7Av1G094570	228864279	228871552	Squamosa promoter-binding protein, putative

TRITD7Av1G094750	229344698	229345696	Lectin receptor kinase
TRITD7Av1G094760	229346051	229346902	Receptor-like kinase
TRITD7Av1G094810	229552507	229555689	Lectin receptor kinase
TRITD7Av1G094840	229614419	229617694	E3 ubiquitin-protein ligase ORTHRUS 2 G
TRITD7Av1G094870	229646021	229648558	Pentatricopeptide repeat-containing protein
TRITD7Av1G094880	229660022	229661149	Sulfotransferase
TRITD7Av1G095000	229788701	229789864	Kinase-like
TRITD7Av1G095050	229874465	229877971	electron transporter, putative (Protein of unknown function, DUF547)
TRITD7Av1G095060	229877537	229879437	Mitochondrial transcription termination factor family protein
TRITD7Av1G095070	229880908	229887454	Dihydroflavonol-4-reductase
TRITD7Av1G095080	229883218	229886972	Laminin-like protein epi-1
TRITD7Av1G095110	229913559	229917910	Chaperone protein dnaJ, putative
TRITD7Av1G095170	230031458	230037123	Peroxidase

TRITD7Av1G095190	**230038058**	**230039686**	**(RAP Annotation release2) Galactose-binding like domain containing protein**
TRITD7Av1G095200	**230040373**	**230047509**	**Glycine-rich domain-containing protein 2 G**
TRITD7Av1G095410	230621524	230626503	RanBP2-type zinc finger protein
TRITD2Av1G018310	35012253	35013821	Glycosyltransferase
TRITD2Av1G018390	35094299	35095598	P53/DNA damage-regulated protein G
TRITD2Av1G018410	35100996	35102480	Glycosyltransferase
TRITD2Av1G018420	35115384	35115935	transmembrane protein, putative (DUF594)
TRITD2Av1G018430	35116075	35117376	transmembrane protein, putative (DUF594)
TRITD2Av1G018440	35118886	35127255	NAD-dependent protein deacetylase
TRITD2Av1G018480	35174326	35174931	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase G
TRITD2Av1G018490	35208236	35209419	Strictosidine synthase

TRITD2Av1G018540	35269369	35290693	Cytochrome P450 family protein, expressed
TRITD2Av1G018620	35292447	35292656	Rho guanine nucleotide exchange factor 2 G
TRITD2Av1G018640	35337809	35341118	O-glucosyltransferase rumi-like protein
TRITD2Av1G018670	35477061	35478293	F-box family protein
TRITD2Av1G018720	35601493	35601954	G-type lectin S-receptor-like Serine/Threonine-kinase G
TRITD2Av1G018730	35606371	35608318	Protein trichome birefringence
TRITD2Av1G018770	35646272	35647252	Dof zinc finger protein
TRITD2Av1G018800	35698472	35702398	Kinase family protein
TRITD2Av1G018810	35703004	35705988	Peptide transporter family protein G
TRITD2Av1G018820	35709157	35714150	2-oxoglutarate dehydrogenase E1 component family protein
TRITD2Av1G018830	35717410	35719557	Telomere repeat-binding factor like-protein

TRITD2Av1G018880	35785954	35790852	methyl-coenzyme M reductase II subunit gamma, putative (DUF3741)
TRITD2Av1G018890	35792786	35793992	Pectate lyase
TRITD2Av1G018900	35796133	35796464	LITAF-domain-containing protein
TRITD2Av1G018910	35818199	35846918	Actin cross-linking protein, putative (DUF569)
TRITD2Av1G018940	35996729	35997151	DnaJ domain containing protein, expressed
TRITD2Av1G018950	35997197	35999320	DNAJ heat shock N-terminal domain-containing protein-like
TRITD2Av1G018990	36059240	36059746	Actin cross-linking
TRITD2Av1G019050	**36289642**	**36293816**	**Kinase interacting (KIP1-like) family protein**
TRITD2Av1G019060	36313105	36314157	Expansin protein
TRITD2Av1G019080	36327882	36328444	Ribosome production factor 2-like protein G
TRITD2Av1G019100	36350313	36351698	UDP-glycosyltransferase
TRITD2Av1G019110	36356017	36359578	Expansin protein
TRITD2Av1G019120	36363569	36364523	Expansin protein

TRITD2Av1G019130	36369506	36381904	Low affinity potassium transport system protein kup G
TRITD2Av1G019170	36444868	36446845	Transcription factor protein
TRITD2Av1G019180	36452268	36455012	Protein kinase-like protein
TRITD2Av1G019200	36522474	36523052	RING/U-box superfamily protein
TRITD2Av1G019210	36526161	36535831	GRF zinc finger family protein, expressed TE?
TRITD2Av1G019220	36537714	36543948	E3 ubiquitin-protein ligase SINA-like 10 G
TRITD2Av1G019230	36558965	36559923	40S ribosomal protein S7 G
TRITD2Av1G019240	36562771	36563358	
TRITD2Av1G019250	36567124	36570283	Pseudo-response regulator
TRITD2Av1G019290	36691780	36696573	Hippocampus abundant transcript-like protein 1 G
TRITD2Av1G019320	36744246	36750772	Sporulation protein RMD1
TRITD2Av1G019330	36751677	36755494	Ascorbate peroxidase
TRITD2Av1G019340	36756356	36758653	Arogenate dehydratase
TRITD2Av1G019400	36873797	36874456	Tapetum determinant 1 G
TRITD2Av1G019430	36890646	36892887	UvrABC system protein C

TRITD2Av1G019440	36926020	36927249	Serine/threonine-protein kinase ATM G
TRITD2Av1G019460	36959198	36960913	Glycosyltransferases
TRITD2Av1G019470	37074527	37076234	F-box family protein
TRITD2Av1G019520	37293639	37384029	Elongation factor 1-alpha
TRITD2Av1G019530	37376072	37376969	Powder tolerance-related protein G
TRITD2Av1G019540	37386189	37387261	Yellow stripe-like transporter 12
TRITD2Av1G019550	37386220	37386411	Carbon storage regulator homolog G
TRITD2Av1G019570	37397770	37408859	MLO-like protein
TRITD2Av1G019650	37586613	37587843	Peroxidase

Table S21: Genes included in the confidence interval of QTLs on chromosomes 6B for plant growth habit trait detected with Blink model in DUS dataset. The genes associated with the peaks are highlighted.

Gene stable ID	**Gene start (bp)**	**Gene end (bp)**	**Gene description**
TRITD6Bv1G204950	637242440	637243795	RING finger protein
TRITD6Bv1G205020	637517818	637520702	Ankyrin repeat-containing protein, putative
TRITD6Bv1G205050	637551481	637551831	arginine N-methyltransferase, putative (DUF688) G
TRITD6Bv1G205060	637553627	637554551	Lectin receptor kinase
TRITD6Bv1G205080	637557890	637559911	Lectin receptor kinase
TRITD6Bv1G205110	637566761	637568794	Lectin receptor kinase
TRITD6Bv1G205130	637635460	637635843	60S ribosomal protein L10a-3 G
TRITD6Bv1G205140	637667900	637668467	Uroporphyrinogen decarboxylase G
TRITD6Bv1G205180	637728905	637731898	Zinc finger CCHC domain-containing protein 12

TRITD6Bv1G205220	637862983	637864299	UDP-D-glucuronate 4-epimerase
TRITD6Bv1G205260	637954722	637957680	F-box protein
TRITD6Bv1G205270	637975632	637987536	Kinase
TRITD6Bv1G205280	**637988454**	**637992747**	**Autophagy-related protein 18**
TRITD6Bv1G205440	**638501886**	**638502842**	**Tyrosine-protein phosphatase CpsB**
TRITD6Bv1G205450	638509062	638512673	Zinc finger CCCH-type with G patch domain protein
TRITD6Bv1G205470	638609480	638609755	plastidic GLC translocator G
TRITD6Bv1G205500	638625337	638628174	Serine incorporator
TRITD6Bv1G205520	638628865	638630800	Pheophorbide a oxygenase, chloroplastic
TRITD6Bv1G205610	638903580	638911872	Isopentenyl-diphosphate delta-isomerase
TRITD6Bv1G205620	638915048	638915536	Protease inhibitor/seed storage/lipid transfer protein family protein G
TRITD6Bv1G205670	639067763	639070684	Disease resistance protein RPM1
TRITD6Bv1G205700	639128882	639143753	Vacuolar protein-sorting protein 33
TRITD6Bv1G205710	639185314	639271835	D-lactate dehydrogenase, putative
TRITD6Bv1G205730	639282042	639282593	Myb/SANT-like DNA-binding domain protein G
TRITD6Bv1G205750	639292866	639303581	Clustered mitochondria protein
TRITD6Bv1G205870	639487878	639488186	Membrane steroid-binding protein
TRITD6Bv1G205890	639536560	639541221	Embryogenesis transmembrane protein-like
TRITD6Bv1G205910	639545539	639549009	ATP-dependent DNA helicase TE?
TRITD6Bv1G205920	639549614	639553062	Snf1-related kinase interactor 1, putative G

TRITD6Bv1G205930	639596249	639599379	RING/U-box superfamily protein

Table S22: Genes included in the confidence interval of QTLs on chromosomes 7A for ear density trait detected with Blink model in DUS dataset. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD7Av1G024980	48253369	48254715	Pectate lyase
TRITD7Av1G025010	48365580	48366025	Ubiquitin-conjugating enzyme 23 isoform 2 G
TRITD7Av1G025110	48732082	48732882	Glycosyltransferase G
TRITD7Av1G025200	49035114	49041714	Elongation factor 4
TRITD7Av1G025250	49243760	49244887	1-aminocyclopropane-1-carboxylate oxidase homolog 2
TRITD7Av1G025340	49929392	49930359	Auxin repressed/dormancy associated protein
TRITD7Av1G025350	49934297	49936115	Ankyrin repeat domain-containing protein 17 G
TRITD7Av1G025430	50200894	50202774	Polygalacturonase-1 non-catalytic beta subunit
TRITD7Av1G025440	50226204	50227788	WAT1-related protein
TRITD7Av1G025520	50455750	50459120	B3 domain-containing protein
TRITD7Av1G025540	50467425	50470754	Pentatricopeptide repeat-containing protein, putative
TRITD7Av1G025550	50472524	50476991	ATP-dependent 6-phosphofructokinase
TRITD7Av1G025650	50791968	50792288	Late embryogenesis abundant protein group 3 protein G

Table S23: Genes included in the confidence interval of QTLs on chromosomes 2B for culm glaucosity of the neck trait detected with Blink model in DUS dataset. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD2Bv1G003900	8254000	8256094	Sulfotransferase
TRITD2Bv1G003910	8257268	8283280	Cytochrome P450 family protein, expressed
TRITD2Bv1G003920	8262311	8263963	Cytochrome P450 family protein
TRITD2Bv1G003930	8287061	8288709	Cytochrome P450 family protein
TRITD2Bv1G003960	8330270	8330718	Centrosomal protein of 126 kDa G
TRITD2Bv1G003970	8331068	8332963	FCH domain only protein 1
TRITD2Bv1G004090	8485014	8626487	Cleavage stimulation factor subunit 1 G
TRITD2Bv1G004100	8485470	8486564	SKP1-like protein 4 G
TRITD2Bv1G004140	8604751	8606397	Cytochrome P450 family protein, expressed
TRITD2Bv1G004160	8627534	8630564	Fructokinase-2
TRITD2Bv1G004200	8658981	8660177	Cytochrome P450
TRITD2Bv1G004380	8743141	8752700	Cytochrome P450 family protein
TRITD2Bv1G004410	8785629	8921809	Cytochrome P450 family protein
TRITD2Bv1G004420	8851980	8852348	Defensin
TRITD2Bv1G004470	9015821	9018300	Cytochrome P450 family protein
TRITD2Bv1G004480	9039417	9041491	Cytochrome P450 family protein
TRITD2Bv1G004510	9127254	9129725	Cytochrome P450
TRITD2Bv1G004520	9134824	9136109	Chalcone synthase
TRITD2Bv1G004560	9165804	9166822	Heptahelical transmembrane protein 4
TRITD2Bv1G004570	9184903	9186138	Alpha/beta-Hydrolases superfamily protein
TRITD2Bv1G004620	9251291	9251722	Cytochrome P450
TRITD2Bv1G004640	9359558	9360837	Chalcone synthase
TRITD2Bv1G004650	9459558	9460576	Heptahelical transmembrane protein 4
TRITD2Bv1G004660	9478515	9479750	Alpha/beta-Hydrolases superfamily protein

TRITD2Bv1G004670	9533137	9535194	Cytochrome P450
TRITD2Bv1G004680	9631850	9633129	Chalcone synthase
TRITD2Bv1G004700	9664423	9665187	GRF zinc finger family protein, expressed TE?
TRITD2Bv1G004710	9728696	9729975	Chalcone synthase
TRITD2Bv1G004720	9766412	9767625	Chalcone synthase
TRITD2Bv1G004730	9799101	9799865	GRF zinc finger family protein, expressed TE?
TRITD2Bv1G004740	9812887	9813752	Cytochrome P450
TRITD2Bv1G004770	9823461	9824756	Chalcone synthase
TRITD2Bv1G004780	9835573	9836028	Alpha/beta-Hydrolases superfamily protein G
TRITD2Bv1G004790	9845315	9857366	O-acyltransferase WSD1
TRITD2Bv1G004830	9964967	9968044	Ankyrin repeat-containing protein
TRITD2Bv1G004840	9970677	9993238	Protein DA1-related 1
TRITD2Bv1G004850	9973519	9977697	Protein kinase
TRITD2Bv1G004860	9982824	9985598	NBS-LRR disease resistance protein, putative, expressed
TRITD2Bv1G004870	10001468	10003936	Serine-rich 25 kDa antigen protein G
TRITD2Bv1G004880	10015129	10016280	Protein PLANT CADMIUM RESISTANCE 2
TRITD2Bv1G004890	10022299	10023232	Protein PLANT CADMIUM RESISTANCE 2
TRITD2Bv1G004900	10038506	10039691	Protein PLANT CADMIUM RESISTANCE 2
TRITD2Bv1G004920	10062924	10065549	Serine/threonine-protein kinase
TRITD2Bv1G004940	10077655	10084624	Protein kinase
TRITD2Bv1G004950	10101254	10105520	Ankyrin repeat protein family-like protein
TRITD2Bv1G004960	10240547	10248379	Protein kinase
TRITD2Bv1G005030	10287867	10289297	NB-ARC domain-containing disease resistance protein

TRITD2Bv1G005050	10349156	10354893	Meiosis-specific protein PAIR3 G
TRITD2Bv1G005060	10403852	10407026	Embryogenesis transmembrane protein-like
TRITD2Bv1G005080	10420483	10425070	Pectate lyase
TRITD2Bv1G005090	10450743	10452200	NB-ARC domain-containing disease resistance protein
TRITD2Bv1G005130	10461824	10462729	NB-ARC domain-containing disease resistance protein
TRITD2Bv1G005140	10507519	10508475	NB-ARC domain-containing disease resistance protein
TRITD2Bv1G005160	10514560	10516050	NB-ARC domain-containing disease resistance protein
TRITD2Bv1G005170	10520775	10524684	Apyrase
TRITD2Bv1G005210	10578069	10579673	NB-ARC domain-containing disease resistance protein G
TRITD2Bv1G005250	10599654	10601111	NB-ARC domain-containing disease resistance protein
TRITD2Bv1G005260	10672965	10673579	Retarded palea 1 protein
TRITD2Bv1G005290	10785190	10786587	NB-ARC domain-containing disease resistance protein

Table S24: Genes included in the confidence interval of QTLs on chromosomes 1A for ear colour trait detected with Blink model in DUS dataset. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD1Av1G000090	130446	131722	Amino acid transporter
TRITD1Av1G000270	724765	733590	Polyphenol oxidase
TRITD1Av1G000280	752684	754056	Histone H3
TRITD1Av1G000360	981483	985759	receptor kinase 1
TRITD1Av1G000380	991776	993030	Werner Syndrome-like exonuclease
TRITD1Av1G000400	999821	1001495	
TRITD1Av1G000450	1095301	1106930	Paired amphipathic helix protein Sin3
TRITD1Av1G000460	1116308	1117751	Pre-mRNA-splicing factor SYF1 G
TRITD1Av1G000480	1156772	1162539	Phospholipid-transporting ATPase
TRITD1Av1G000490	1168983	1172033	Cytochrome P450
TRITD1Av1G000500	1175277	1177769	Endoribonuclease E-like protein G
TRITD1Av1G000530	1179103	1181600	Receptor-like kinase
TRITD1Av1G000550	1197264	1198501	Nucleic acid-binding, OB-fold
TRITD1Av1G000570	1210389	1216942	vascular related NAC-domain protein 1 G
TRITD1Av1G000580	1219368	1227585	RING-finger ubiquitin ligase
TRITD1Av1G000590	1241353	1242498	50S ribosomal protein L28
TRITD1Av1G000850	1990406	1990702	30S ribosomal protein S14 type Z G
TRITD1Av1G000900	2064766	2143117	lectin protein kinase family protein G
TRITD1Av1G000910	2075070	2075489	Senescence-associated protein, putative G
TRITD1Av1G000950	2095681	2095977	NAD(P)-binding Rossmann-fold superfamily protein G
TRITD1Av1G000960	2102522	2102908	lectin protein kinase family protein G
TRITD1Av1G000970	2102545	2102960	Formin-like protein G
TRITD1Av1G000990	2112078	2112338	TRICHOME BIREFRINGENCE-LIKE 26 G
TRITD1Av1G001010	2123377	2123673	ATP synthase gamma chain G

TRITD1Av1G001030	2131632	2131928	Mediator of RNA polymerase II transcription subunit 12 G
TRITD1Av1G001060	2221027	2221404	Protein TAR1 G
TRITD1Av1G001070	2231454	2231831	lectin protein kinase family protein G
TRITD1Av1G001090	2246308	2251377	Polyketide cyclase/dehydrase and lipid transport superfamily protein G
TRITD1Av1G001100	2250939	2251354	myosin 2 G
TRITD1Av1G001120	2278926	2279222	ATP synthase gamma chain G
TRITD1Av1G001140	2295761	2296057	Coiled-coil domain-containing protein G
TRITD1Av1G001160	2338627	2338887	tRNA (cytidine(34)-2'-O)-methyltransferase G

Table S25: Genes included in the confidence interval of QTLs on chromosomes 1A for awn colour trait detected with Blink model in DUS dataset. The genes associated with the peaks are highlighted.

Gene stable ID	Gene start (bp)	Gene end (bp)	Gene description
TRITD1Av1G000090	130446	131722	Amino acid transporter
TRITD1Av1G000270	724765	733590	Polyphenol oxidase
TRITD1Av1G000280	752684	754056	Histone H3
TRITD1Av1G000360	981483	985759	receptor kinase 1
TRITD1Av1G000380	991776	993030	Werner Syndrome-like exonuclease
TRITD1Av1G000400	999821	1001495	
TRITD1Av1G000450	1095301	1106930	Paired amphipathic helix protein Sin3
TRITD1Av1G000460	1116308	1117751	Pre-mRNA-splicing factor SYF1 G
TRITD1Av1G000480	1156772	1162539	Phospholipid-transporting ATPase
TRITD1Av1G000490	1168983	1172033	Cytochrome P450
TRITD1Av1G000500	1175277	1177769	Endoribonuclease E-like protein G
TRITD1Av1G000530	1179103	1181600	Receptor-like kinase
TRITD1Av1G000550	1197264	1198501	Nucleic acid-binding, OB-fold
TRITD1Av1G000570	1210389	1216942	vascular related NAC-domain protein 1 G
TRITD1Av1G000580	1219368	1227585	RING-finger ubiquitin ligase
TRITD1Av1G000590	1241353	1242498	50S ribosomal protein L28
TRITD1Av1G000850	1990406	1990702	30S ribosomal protein S14 type Z G
TRITD1Av1G000900	2064766	2143117	lectin protein kinase family protein G

TRITD1Av1G000910	2075070	2075489	Senescence-associated protein, putative G
TRITD1Av1G000950	2095681	2095977	NAD(P)-binding Rossmann-fold superfamily protein G
TRITD1Av1G000960	2102522	2102908	lectin protein kinase family protein G
TRITD1Av1G000970	2102545	2102960	Formin-like protein G
TRITD1Av1G000990	2112078	2112338	TRICHOME BIREFRINGENCE-LIKE 26 G
TRITD1Av1G001010	2123377	2123673	ATP synthase gamma chain G
TRITD1Av1G001030	2131632	2131928	Mediator of RNA polymerase II transcription subunit 12 G
TRITD1Av1G001060	2221027	2221404	Protein TAR1 G
TRITD1Av1G001070	2231454	2231831	lectin protein kinase family protein G
TRITD1Av1G001090	2246308	2251377	Polyketide cyclase/dehydrase and lipid transport superfamily protein G
TRITD1Av1G001100	2250939	2251354	myosin 2 G
TRITD1Av1G001120	2278926	2279222	ATP synthase gamma chain G
TRITD1Av1G001140	2295761	2296057	Coiled-coil domain-containing protein G
TRITD1Av1G001160	2338627	2338887	tRNA (cytidine(34)-2'-O)-methyltransferase G

5 Characterization and fine mapping of the *sbm2* QTL for resistance to SBCMV

5.1 Introduction

5.1.1 Characterization of Soil Borne Cereal Mosaic Virus (SBCMV) resistance in durum and bread wheat

Agriculture will face major challenges during the next decades because of climate change and increase of population. Disease resistance is one of the principal topics that is currently being explored in different species and cultivars, in order to register resilient varieties to principal pests in common market. One of the most severe disease for durum and bread wheat, but grain cereals in general, is represented by the Soil-borne Viruses, which cause sever yield losses (Kühne, 2009).
In particular, Soil Borne Cereal Mosaic Virus (SBCMV) belongs to the *Furovirus* genus and, like other Soil Borne Viruses, is vectorized by the soil fungus *Polymixa graminis* (Budge et al., 2007; Maccaferri et al., 2011; Ratti et al., 2004; Rubies., 2003; Brakke & Langenberg, 1988.) where viral particles are protected by the environment and spores may remain silent even for different years. This inter-cellular parasite develops inside the roots of wheat plants: when a susceptible cultivar variety is reached and environmental conditions are optimal, viral zoospores are released from resting spores which incorporates the pathogen, entering the plant root hair cells (Campbell, 1996; Kühne, 2009). In infected plants, viral zoospores diffuse from the infected roots, reaching the aerial parts of the plant where it causes range of different symptoms that causes severe yield losses during the growing season. The visual symptoms go from different chlorosis diffusion in the leaf surface, differentially distributed with different degrees of severity. The most severe disease symptoms cause a strong depletion of yield production, resulting in up to 70% of yield losses which is worsening due to climate change.
The genus *Furovirus* contains 5 different species differentially distributed (Bass et al., 2006). Namely, SBCMV represents the European race (France, Germany, Italy, Denmark, Poland and UK), Soil Borne Wheat Mosaic Virus (SBWMV) was first identified in North America causing severe yield losses but is spread also in southern Germany and Africa (Bass et al., 2006; Ratti et al., 2004), and Chinese wheat mosaic virus (CWMV) distributed in different part of Asia (Budge et al., 2007). The SBCMV genome is divided in two RNA molecules: RNA1 codifies for a 37 kD polymerase described as the movement protein, RNA2 codifies for a cysteine rich repeat protein (unknown function) and for the viral cap protein (Kühne, 2009). The latter is the only molecule transferred. Furthermore, as regards SBWMV which shares a very similar genomic structure to SBCMV, RNA2 was found to be deleted spontaneously in case of dilatated presence of the virus inside the host (Chen et al., 2014) (Figure 43).

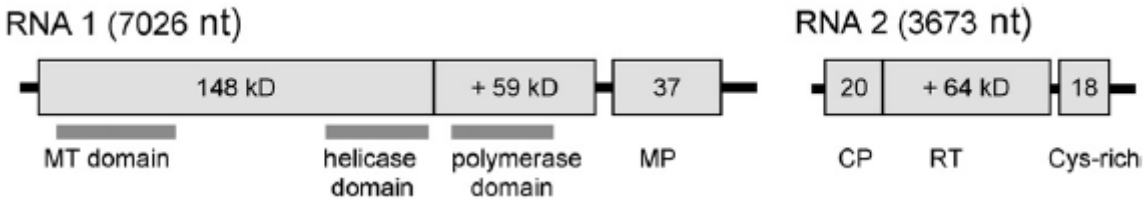


Figure 43: RNA1 and RNA2 structure in soil borne viruses. RNA1 codifies for the movement protein (MP), the polymerase domain and the methyltransferase. RNA2 codifies for the coat protein, the readthrough protein and the cysteine rich domain (Kühne, 2009).

Bread and durum wheat are strongly affected by SBCMV, causing yield losses between 50% and 70% in France and Italy (Budge et al., 2007; Rubies-Autonell et al., 2003). Symptoms presence can be evaluated by visual score analysis on leaves and on plant growth habit, however the severity of mosaic symptoms can depend on the genotype (resistant or susceptible), the aggressiveness of viral strains and environmental conditions (Budge et al., 2007; Kanyuka et al., 2004). Nevertheless, the presence of non-homogeneous parts of field could be responsible of false positive regarding resistance of genotypes to SBCMV. In addition to symptoms visual evaluation, it is therefore important, in order to perform genetic studies, to perform molecular evaluation of infection levels using enzyme-linked immunosorbent assay (ELISA) tests or reverse transcription–polymerase chain reaction (RT–PCR) analysis (Kanyuka et al., 2004; Ratti et al., 2004). Because of the persistent nature of SBCMV, the only solution against SBCMV is the development of resistant varieties, monitored for different years on mosaic virus infected fields, in order to select economically valuable varieties resistant to the disease (Budge et al., 2007; Rubies-Autonell et al., 2003). On the other hand, the agronomic practices like crop rotation or delayed sowing are ineffective.

Genetic resistance in common bread wheat was detected in one to three major QTLs (Barbosa et al., 2001; Bass et al., 2006; Dubey et al., 1970; Narasimhamoorthy et al., 2006).

Genetic resistance was detected both in bread and durum wheat in two different QTLs: s*bm1* (Bass et al., 2006) and *sbm2* (Maccaferri et al., 2011a). Resistance in bread wheat was detected in *Triticum aestivum* cv. Cadenza evaluating a double haploid population composed by Avalon (susceptible cultivar) and Cadenza. Resistance was found to be controlled by a single locus based on the segregation ratio, and consists in a mechanism called “translocation resistance” that prevents the infection to spread from the roots to the stem and leaves (Kanyuka et al., 2004; Ordon et al., 2009). The *sbm1* QTL was mapped in long arm of chromosome 5D in Bread wheat, exploiting microsatellites SSR markers where *Xbarc110* and *Xwmc765* were found to be flanking *Sbm1* QTL in a region of approximately 17 cM (Bass et al., 2006). This QTL showed resistance also to SBWMV (Narasimhamoorthy et al., 2006). The *sbm2* QTL was instead mapped on the short arm of chromosome 2B (Bayles et al., 2007), showing that resistance to SBCMV is controlled by different major genes (Maccaferri et al., 2011a). The resistance in durum wheat was ascribed to *sbm2* by the analysis of 181 Recombinant Inbred Lines (RILs) belonging to the segregant population Meridiano (resistant) x Claudio (susceptible) (Maccaferri et al., 2011a). Based on multiple years field trials to evaluate RILs response, a major QTL, namely *QSbm.ubo-2B* (confirmed to be *sbm2*), was found to be responsible for phenotypic variance to SBCMV resistance. The QTL was delimited by SRR *Xwmc661-Xgwm210-Xbarc35,* with DarT marker wPt-2106. The same QTL was detected in 180 individuals from the cross Simeto (susceptible) and Levante (resistant), being responsible for 60-70% of phenotypic diversity.

Previous studies conducted to dissect the genetic basis of SBCMV resistance in durum wheat were based on SSR and DArT markers, massively exploited as “past generation markers”. However, these marker classes present a series of constraints: low throughput (SSRs), density insufficient to achieve

a high mapping resolution, required in particular for fine mapping (SSR and DArT) (Terracciano et al., 2013) and limited informativeness (DArT) (Maccaferri et al., 2012).
As a consequence of the widespread use of SNP molecular marker, different types of genotyping arrays were developed, such as Illumina 90K (Wang et al., 2014a) and Affimetrix SNP Chip 35K and 820K (Allen et al., 2017) to perform genome association studies and QTL mapping. In addition, the development of the consensus map for tetraploid wheat, which anchored approximately 90K SNP (Maccaferri et al., 2015), and the sequencing of *Triticum turgidum* cv Svevo genome (Maccaferri et al., 2019b) allowed to physically map the SNPs arrays. These were major advantages for genome wide association studies (GWAS) and candidate gene evaluation. Genome-wide association studies (GWAS) coupled with dense SNPs arrays allowed to efficiently exploit large phenotypic datasets to search for new QTLs and confirm the presence of QTLs already mapped (Korte and Farlow, 2013).
Further research was therefore performed on the durum panel (Maccaferri et al., 2015) and on two experimental population composed by Meridiano (Resistant) x Claudio (Susceptible) and Simeto (Susceptible) x Levante (Resistant) (Maccaferri et al., 2011; Bruschi et al., *unpublished*). Exploiting the SNP90K Consensus Map (Maccaferri et al., 2015), 9 KASP markers were developed and validated on the two segregating population, fine mapping the Sbm2 interval, initially detected by Maccaferri et al (2011; 2012), was performed narrowing down the interval to 3 Mb, from 13Mbp to 16Mbp on chromosome 2BS. In parallel, a GWAS was performed on the durum panel phenotyped for resistance to SBCMV, representing a wide number of modern durum varieties, confirming the presence of the *sbm2* QTL on chr2B and validation of the previously designed KASP markers (Bruschi et al 2022 unpublished).

5.2 Material and methods

5.2.1 Background analysis and plant material

The starting point of this work was background material and analysis already performed in Unibo, as a consequence of the research performed by Maccaferri and Bruschi et al (Maccaferri et al., 2011; 2012; Bruschi et al., *unpublished*). Briefly, a population composed of 181 RILs was developed by Produttori sementi Bologna Spa (Argelato, Bologna) by crossing elite durum varieties Meridiano (M) and Claudio (C); the population was advanced until F_7:F_8 stage by single seed descent (Maccaferri et al., 2011). This population was useful to perform a backcross by the introgression of Meridiano in Meridiano x Claudio RILs (MxMC) producing 2500 lines in F_2. These lines were analyzed by PCR with KASP markers developed by Illumina 90K SNP Chip (Bruschi et al., *unpublished*) flanking the *sbm2* confidence interval KUBO 9 – wpt2106 (DaRT marker) and KUBO 13. 600 recombinants were detected and advanced to homozygous stage in greenhouse and used for further analysis. Meridiano variety is resistant to SBCMV, medium early heading under Mediterranean conditions and high yield production (pedigree Simeto/WB881/Duilio/F21). Claudio variety is a medium late heading, high yield production, medium susceptible to SBCMV but good resistance to oidium, fusarium, septoria and leaf rust, medium resistance to cold and tillering (Maccaferri et al., 2011; Rubies-Autonell et al., 2009).
Modern durum varieties were sown during 2020/2021 growing season in order to perform a transcriptomic experiment using RNAseq technology. Resistant and susceptible varieties were sown, namely: Neodur (R), Meridiano (R), Svevo (R), Levante (R), Simeto (S), Ciccio (S), Claudio (S) and Altar 84 (S).

5.2.2 Field trials

Field trials were conducted in Cadriano (BO) (44°35' N 11°27' E) in a field with homogeneous presence of SBCMV infection as a result of continuous growth of durum Grazia cultivar (susceptible to SBCMV). 600 RILs detected through MAS in MxMC population were sown during early November 2017 in SBCMV field, two complete block replicates sowed randomly at a seed density of 350/400 seed/m2. Each plot was composed of two rows 1.20 m long divided from the following plot by 1m. In addition, Svevo, Meridiano Ciccio, Grazia, Simeto, Levante and Claudio parental lines were sown in regular intervals together with RILs in order to control the rate of infection and its distribution.
During November 2018, after the second step of fine mapping occurred, 17 RILS MxMC recombinant for the restricted interval were detected and sown in SBCMV field. In addition, 39 MxMC RILs were sown in order to repeat the phenotypic test. Finally, after a third evaluation with molecular markers, 6 critical recombinants were sown in randomized block design with 3 replicates during 2020/2021 growing season and evaluated in Cadriano nursery.
The RNAseq experiment field trial was performed in randomized block design with 3 replicates. Roots and leaves were sampled for RNA extraction from a bulk of 3 plants, sampling occurred on 04-02-2022, 18-02-2022, 04-03-2022 and 11-03-2022. The samples were instantly frozen in liquid nitrogen and kept at -80°C.

5.2.3 Phenotypic evaluation

SBCMV presence was detected on RILs evaluated in field trials by Visual score (VS) and ELISA test. VS reports the symptom severity (SS) with a score from 0 to 5, where 0 reports resistance whilst 5 is susceptible. The scale was reported by Vallega and Autonell (Vallega, 1985) as following: 0-1.5 = no or slight symptoms, 1.51-2.5 = mild mottling and stunting, 2.51-3.5 = mottling and stunting, 3.51-4.5 = severe mottling and stunting, 4.51-5.0 = plants killed by virus. In particular, it is worth pointing out that the scale developed by Vallega and Autonell reports SS score until 4, but the scale has been extended up to 5 in order to include plant killed by virus infection and to separate better susceptible and resistant scores. SS was detected on leaves surface in few VS scoring dates every week starting from March to the beginning of April, when the plant starts to show the first symptoms until the end of the virus infection. This period starts from the mid-end of tillering stage (GS25) (Zadoks et al., 1974) until first and second node appearance (GS31 and GS32). As regards 2017/2018 growing season, VS was performed on MxMC RILS in 6 separate dates: 22-03-2018, 28-03-2018, 6-04-2018, 11-04-2018, 16-04-2018, 20-04-2018. On the other hand, during 2018/2019, the VS was performed on MxMC, in 4 separate dates: 11-03-2018, 19-03-2019, 26-03-2019, 2-04-2018. During 2021/2022, MxMC was scored on 4 different dates: 02-03-2022, 08-03-2022, 16-03-2022 and 25-03-2022. During the same year, modern varieties for transcriptomics were sampled and evaluated for SS during the following dates: 04-02-2022, 18-02-2022 and 04-03-2022. These growth stages were between third leaf unfolded (GS13) and tillering complete (GS29), as the aim was to evaluate the progress of the disease from the beginning of infection.
The ELISA test was performed during 2019 on all the material sown in the field trials using double antibody sandwich according to protocol reported by Clark and Adams (1977). Antiserum was prepared using SBCMV particles extracted from Grazia cv. sown in field trials.

5.2.4 KASP marker design and validation

Starting from the KASP markers developed from the Illumina SNP 90K consensus map (Maccaferri et al., 2015) and tested on the durum panel and on the biparental population Svevo x Ciccio and Meridiano x Claudio (Bruschi et al., *unpublished*), the following markers were tested on the MxMC BCF_4 population (Table 26 and Table 27).

Table 26: Illumina 90K SNP Chip ID, KUBO ID and probe sequence with allelic variants in bold

SNP	KUBO ID	Sequence
IWB73347	**KUBO13**	AATCGGTATATAACTTCCGCAGAAAGGTGCATCTCTGACATTCTCAGAAC[A/G]GCA ATGTCCCTGTCAGCCGGATGTGCTCGAGGGTGCGCTTCAGGCCGTAC
IWB11421	**KUBO27**	ATTCGTTGCTGCTCTTATACTTGAGGATTATATCTACTTCAAGTTTCACC[A/G]TTGCT TATTAGCGAGTGATTGTTTATTCTGCTGAAATCATTTGAGCTCCT
IWB8328	**KUBO3**	AAAAGGCCTACTATCCAGGATATTGTCCGTGAAGTGGAGGAACTAGAAGC[T/C]GA TATTGAGAAAATGATGTCAGCACCTTCTCCTGAGTCAAAAGATCTAAC
IWB28973	**KUBO1**	TTGATCTCTATGTTGAATGCGATGAAGCCACCCGTCTTGTTGGTCAGCTG[T/C]AGG CAGCACGACGCCTCCTTCCTTAGCTCAAAGAGGAACCGCAGCTCCAT
IWB29097	**KUBO40**	GACCTATCGAATCATGAGGACCTGGGGCGTTGCACGGGACGATGACTCGT[T/C]TC CTGGGATAAAAGGAGCATCAGCCCCCAGTTTCATTTCATATGAAATTG
IWB35524	**KUBO41**	AATATATGCAGGGCGGAAGCCTTGAGAAATACATCGCAGACGAACCATGTAGACTT GACTGGCCCACATGTTATAAAATCATTCAAGGGATCTGTGACGG[T/C]TTAAATCAC CTTCACAATGCACAGGAAAAAGCAATTTTCCATCTGGACTTGAAGCCTTCGAA
IWB10512	**KUBO9**	CATTGTCGGAACATGAGTTATCTCACTAGGATGCGTTGTTTGCCAACTCT[A/G]TAA CTTATGGTGGTGGTGGATGGCCCAAGGGATGTGGCGCTCAACTCCAT

Table 27: KASP markers converted from 90K SNP chip. Allele 1 and 2 tails are complementary to FAM and HEX probes used in KASP PCR kit. GAAGGTGACCAAGTTCATGCT = Allele 1 tail labelled oligo sequence, added to primer_A at 5'end AAGGTCGGAGTCAACGGATT = Allele 2 tail labelled oligo sequence, added to primer_B at 5' end Primer A and B are allele specific, Primer C is genome specific for genome B (sbm2 QTL).

SNP ID	KASP ID	Primer A	Primer B	Primer C
IWB73347	**KUBO13**	GAAGGTGACCAAGTTCATGCTggctgacagggacattgcT	GAAGGTCGGAGTCAACGGATTggctgacagggacattgcC	gtgcatttcaaaattcattttctc
IWB11421	**KUBO27**	GAAGGTGACCAAGTTCATGCTtaaacaatcactcgctaataagcaaC	GAAGGTCGGAGTCAACGGATTtaaacaatcactcgctaataagcaaT	gccatatgctgtgactgctg
IWB23029	**KUBO29**	GAAGGTGACCAAGTTCATGCTgctggaggaactagaagctgaG	GAAGGTCGGAGTCAACGGATTgctggaggaactagaagctgaT	aacagagcattgcaaaaccta
IWB8328	**KUBO3**	GAAGGTGACCAAGTTCATGCTtgaagtggaggaactagaagcT	GAAGGTCGGAGTCAACGGATTtgaagtggaggaactagaagcC	gagaaggtgctgacatcattttc
IWB28973	**KUBO1**	GAAGGTGACCAAGTTCATGCTcgtcttgttggtcagctgT	GAAGGTCGGAGTCAACGGATTcgtcttgttggtcagctgC	gcagctgtgacacgaacatta
IWB29097	**KUBO40**	GAAGGTGACCAAGTTCATGCTgatgctccttttatcccaggaA	GAAGGTCGGAGTCAACGGATTgatgctccttttatcccaggaG	tcgaatcatgaggacctggg
IWB35524	**KUBO41**	GAAGGTGACCAAGTTCATGCTtgtGcattgtgaaggtgatttaaA	GAAGGTCGGAGTCAACGGATTtgtGcattgtgaaggtgatttaaG	acttgcagacgaaccatgtag
IWB10512	**KUBO9**	GAAGGTGACCAAGTTCATGCTaggatgcgttgtttgccaactctA	GAAGGTCGGAGTCAACGGATTggatgcgttgtttgccaactctG	gccatccaccaccaccataagtta

In order to fine map the *sbm2* confidence region on short arm of 2B chromosome (Bruschi et al., *unpublished*; Maccaferri et al., 2011), these KASP markers were developed from Illumina SNP 90K consensus map (Maccaferri et al., 2015; Liu et al., 2014). Using Polymarker tool (http://Polymarker.tgac.ac.uk), 25 KASP primers were designed within the *sbm2* confidence interval but, after PCR test, only 8 proved to be polymorphic in resistant and susceptible parental lines from Illumina SNP 90K Chip (IWB markers) (Table 27). According to KASP protocol developed by (LGC Genomics) KASP primers were developed as following: primer A and primer B were designed on the varietal SNP, each one presents at 3'-end a specific allelic variant and at 5' end the complementary sequence for two different fluorophore FAM and HEX which emit fluorescence at different wavelength (GAAGGTGACCAAGTTCATGCT for 5'-FAM in primer_A and GAAGGTCGGAGTCAACGGATT for 5'-HEX in primer_B). Primer C is common for primer A and B but it is genome specific and was designed on homeologue SNP at 3'- end. KASP PCR components were reported in Table 28.

Table 28: components in each well plate 96x for PCR reaction with KASP primers.

PCR reaction (well)	
DNA (20ng/µl)	2µl
Primer mix	0.14µl
Master Mix 2X	3µl
H_2O	3µl

Primer mix was composed by Primer A, B and C in order to dilute them to working solution (Table 29).

Table 29: primer mix diluted to working solution for KASP PCR genotyping.

Primer mix	
Primer A (100µM)	12µl
Primer B (100µM)	12µl
Primer C (100µM)	30µl
H2O	46µl

KASP PCR genotyping tests were developed in 96x well optical plates and run in FAST 7500 Bio Analyzer (Applied Biosystem) using the following thermal protocol (Table 30).

Table 30: thermal cycling protocol for KASP PCR using Fast 7500 thermal cycler machine.

Pre-PCR Read	30°C for 1 minute
Step 1 - Polymerase activation step	94°C for 15 minutes
Step 2 - 10 cycles	94°C for 20 seconds 61-55°C for 60 seconds with a temperature decrease of 0.6°C per cycle
Step 3 - 26 cycles	94°C for 20 seconds 55°C for 60 seconds
Post-PCR Read	30°C for 1 minute

Usually, the number of cycles reported in the standard thermal protocol are not enough to separate allelic varieties in a proper way, thus recycling steps are needed composed by temperature reported in Step 3 and Post PCR Read.
Each KASP marker was tested on resistant (Meridiano, Svevo and Levante) and susceptible (Ciccio, Claudio and Simeto) durum wheat varieties to SBCMV. Artificial heterozygotes were created by mixing 1µl DNA of Meridiano and 1ul DNA of Claudio varieties. Each variety and artificial heterozygote were analysed in 3 technical replicates for each KASP marker.
The KASP markers from Illumina Infinium 90K SNP Chip (Table 26 and Table 27) were developed and tested on background material (Bruschi et al., *unpublished*) and some of them were used for haplotype analysis on a durum panel of modern elite varieties.
Further KASP markers were developed from 420K Affimetrix SNP Chip (cfn markers) (Table 31) to fine map the *sbm2* interval on the backcross population MxMC. 11 KASP markers proved to be polymorphic on *sbm2* contrasting parental lines, out of 22 primers designed. Two primers were designed from Illumina SNP 90K Chip for a total number of 13 KASP marker within the *sbm2* confidence interval covering a region of approximately 2 Mb (Table 31 and Table 32).
Among the 11 KASP markers validated, few of them were designed by hand as Polymarker did not produce polymorphic marker. Primer design by hand was produced using MEGA X software (https://www.megasoftware.net) by multiple alignment of marker sequence and A/B genomes of

Triticum turgidum cv. Svevo, *Triticum dicoccoides* cv. Zavitan and *Triticum aestivum* cv. Chinese Spring (Maccaferri et al., 2019; Avni et al., 2017; The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018).

Table 31: Axiom 420K Affimterix SNP markers sequences, converted into KASP markers.

Marker ID	KASP ID	Marker seq
cfn0628661	**KUBO 60**	GTTATATAGTTCACGCACTAGTAAGTAGTAAAAGA[C/T]GAGGAGCTTGAGGATAGTGATTTATAGGTCCGAAG
cfn0628662	**KUBO 61**	AATAACGGATAGGACTGATAACTAGCAAAACACGC[A/G]TAGTTTCTTTTTGCTGTGCTCAATTTGTATACATA
cfn0631347	**KUBO 62**	AGGCCACTCATAAAGGAGAAGAAATTGCCGTGAAT[A/T]CCACGCCGAATACCGTAACCTTCGTAAGGTACGCC
cfn0634141	**KUBO 63**	CAATTTCAGCCTCTGCTCCACTGTTGCTGCAACAC[G/T]CGTTTCCCCTGCTCCCCATGGATGAAAAGCTCACC
cfn0643373	**KUBO 64**	TGCAATCTAGCTAGCTTTTGCTTGGTGAGTACTAG[C/T]GGAGTACTTCTTGGCACTTGGCAGCGAGAGCTTGG
cfn1702104	**KUBO 65**	CCAGTGGAGAATTGGCACGAGGTGGCTGATATGTT[C/T]CTTTTACTTTTTTTTTTGAAAATAGAAGTGATGGA
cfn1744096	**KUBO 66**	ATGGAAATTATTCCGTGCATACATGGATCATGTCA[C/T]AGGTCGATAAGTTGAAGGCCGGCGTGAGGGAGCGG
cfn0631343	**KUBO 68**	CCGTGCAACCAGCAACGCGCTTACCTCCATTATTA[A/G]AAAGGTTGCAAGGAGCTGCTGGACCAGCAAGCAAC
cfn0638744	**KUBO 69**	GCCGCCATGAAAATACCCAAACACGCGCATGCAAG[A/G]GACAGGAACAGGAGAAGACATCGGGCACCCTAACA
cfn0643378	**KUBO 70**	TGAGTGTTAAAAAGGTGGAGTCTCTCCGCACCGTC[G/T]CCATGGAACAGGTAACTTCAAAAACAAAATATCAG
cfn0643594	**KUBO 71**	ATATTGCTGGATAAGAGCATGACGCCTAAAATCGC[C/G]GATCTTGGTTTGTCCAAACTTGTTTCTTCGACATT
cfn1740996	**KUBO 73**	AAAACAGTACAGAAGCACCTTCCTAGGCCTCATGC[C/T]TCCAGCTGCAGGTAGTAGTGTGAAAGAGATAGATG
cfn0643379	**KUBO 74**	GACGAGAAAACAAATAGAAAAATAAAATTATTCAA[A/G]CCAAGAAATGCGCACAGGAAGGCACGTAGTGCGTC
cfn0643388	**KUBO 75**	TGTGATGGACATGAGCAAGCTTCTCAAGTCATTCA[A/C]ATGTGACATTGCGTGCCGCATCGTGTCGGGAGAAT
cfn1729998	**KUBO 76**	CGATTTCATCGGCCCCGACAACCCTCCCCCTGATG[A/G]GAGTATTATGACTCCATTGCCACAAGCTTTTTCTA
IWB40895	**KUBO 80**	CATGATTACCCTCTGGGAAAGCACATGTAGCAACATCCCATACTGATGTA[T/G]CCCATATGTCATCAATTATAATAATATACCTTTTATCCTGTAGATGCTCG
cfn0627610	**KUBO 81**	AGTGTGCTGTGGAGGGCCCGGAAGGTACGGAATGA[A/G]CTCGACTGTTCCATGCGGCGATGCAGCCGCGACGA
cfn0643155	**KUBO 82**	CGTCCCAGAAGCATCAAAAACATTCGCCCACCACT[C/T]ATCAATTCTTTTGAGCGCAGTGTAAGGAACAACTC

Table 32: KASP markers validated to be polymorphic between resistant and susceptible durum wheat varieties. 11 markers were converted from 420K Affimetrix SNP Chip Array (cfn code) and 2 markers from SNP 90K Chip (IWB code). The HEX and FAM tails are reported in capital letters in primer A and B, (GAAGGTGACCAAGTTCATGCT for 5'-FAM in primer_A and GAAGGTCGGAGTCAACGGATT for 5'-HEX in primer_B).

SNP code	KASP ID	Primer A	Primer B	Primer C
cfn0634141	**KUBO 63**	GAAGGTGACCAAGTTCATGCTGGgGagcaggggaaacgC	GAAGGTCGGAGTCAACGGATTGGgGagcaggggaaacgA	cacagaccaacactgcgA
cfn0643373	**KUBO 64**	GAAGGTGACCAAGTTCATGCTgctttgcttggtgAgtacTAgC	GAAGGTCGGAGTCAACGGATTgctttgcttggtgAgtacTAgT	tcgctgccaagtGccaagA
cfn1702104	**KUBO 65**	GAAGGTGACCAAGTTCATGCTcacgAggtggcTgatatgttC	GAAGGTCGGAGTCAACGGATTcacgAggtggcTgatatgttT	cccctcTccatcactctctAtttT
cfn0631343	**KUBO 68**	GAAGGTGACCAAGTTCATGCTcgcgcttacctccattattaa	GAAGGTCGGAGTCAACGGATTcgcgcttacctccattattag	ggcgaagacgagttgctct
cfn0643594	**KUBO 72**	GAAGGTGACCAAGTTCATGCTgcatgacgcctaaaatcgcg	GAAGGTCGGAGTCAACGGATTgcatgacgcctaaaatcgcc	cgaagcaagcctaaagtatgca
cfn1740996	**KUBO 73**	GAAGGTGACCAAGTTCATGCTAcactactaccTgcagctGgaG	GAAGGTCGGAGTCAACGGATTAcactactaccTgcagctGgaA	tagctgtcagattaatgagacaagg
cfn0643379	**KUBO 74**	GAAGGTGACCAAGTTCATGCTcctgtgcgcatttcttggT	GAAGGTCGGAGTCAACGGATTcctgtgcgcatttcttggC	agattatgggttgggaggga
cfn0643388	**KUBO 75**	GAAGGTGACCAAGTTCATGCTgcaagcttctcaagtcattcaA	GAAGGTCGGAGTCAACGGATTgcaagcttctcaagtcattcaC	tattctcccgacacgatgcg
cfn1729998	**KUBO 76**	GAAGGTGACCAAGTTCATGCTgtggcaatggagtcataatactct	GAAGGTCGGAGTCAACGGATTgtggcaatggagtcataatactcc	gcaattgcactcgggggtta
IWB40895	**KUBO 80**	GAAGGTGACCAAGTTCATGCTgtagcaacatcccatactgatgtat	GAAGGTCGGAGTCAACGGATTgtagcaacatcccatactgatgtag	gaaataactgtgtgcagcca
cfn0627610	**KUBO 81**	GAAGGTGACCAAGTTCATGCTCGCATGGAACAGTCGAGC	GAAGGTCGGAGTCAACGGATTcgcatggaacagtcgagt	tggaaggtggttgacgttg
cfn0643155	**KUBO 82**	GAAGGTGACCAAGTTCATGCTcactgcgctcaaaagaattgatG	GAAGGTCGGAGTCAACGGATTcactgcgctcaaaagaattgatA	ccaatttccccatgcgtc

The KASP makers from 420K Affimetrix SNP Chip were tested on 30 RILS MxC population recombinants in *sbm2* confidence interval between KUBO 9 and KUBO 13 and the recombinant bins were confirmed (data not shown) (Bruschi et al, *unpublished*).
600 MxMC RILs were genotyped with the following KASP markers: KUBO 63, KUBO 64, KUBO 65, KUBO 68, KUBO 72, KUBO 73, KUBO 74, KUBO 75 during 2018. Each KASP marker was analyzed also with parental lines Meridiano, Claudio, Svevo, Ciccio, Simeto and Levante to confirm allelic separation and with an artificial Heterozygote created merging equal part of resistant and susceptible cultivar.

5.2.5 DNA extraction

MxMC 600 RILs were sown in greenhouse, 4 seeds per line. At first leaf stage, leaf samples were collected and located in the lyophilizator for three days under vacuum and at -50 °C. Subsequently, leaf samples were grinded with Tissuelyser (QIAGEN) obtaining approximately 100mg of leaf powder that was used for DNA extraction. DNA was extracted following CTAB protocol (Doyle and Doyle, 1987), with few adaptations. For example, two steps of chloroform purifications were applied in order to have a better purification of the DNA.
After the extraction, DNA was quantified with biophotometer (QIAGEN), retaining samples with 260/280 and 260/230 ratios higher than 2.

5.2.6 RNA extraction and RNAseq experimental layout

Root samples from varieties harvested from the field on 4-02-2021, 18-02-2021 and 4-03-2021 were collected and grinded with liquid nitrogen. Total RNA was extracted starting from about 100mg of root samples using the RNA extraction plant kit protocol (Macherey Nagel). The RNA from each sampled variety was extracted and quantified using Nanodrop, retaining only samples with quality ratio higher than 2. RNA samples were diluted at the same concentration in order to prepare the samples for RNAseq.
As regards the RNAseq experiment, Svevo and Ciccio were used as single varieties, then resistant and susceptible bulks were created from other varieties, namely: Neodur, Meridiano and Levante were included in the resistant bulk (RB), Simeto, Claudio and Altar-84 were included in the susceptible bulk (SB). All the samples send to be sequenced were represented in 3 biological replicates, regarding the bulks the merging of varieties between different replicates occurred including the same replicates of each sample.

5.2.7 RNAseq analysis

RNA samples were sequence at IGA Tech company (Udine, Italy) using Illumina technology with 150 bp paired end reads. RNAseq libraries were created following for each sample, producing 20 million pared reads for single varieties (Svevo and Ciccio) and 30 million for resistant and susceptible bulks. RNA reads were checked for quality control using Fastqc (“Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data,” n.d.) with default options. RNAseq reads were trimmed using TrimGalore software (“Babraham Bioinformatics - Trim Galore!,” n.d.), removing adapters used in the Illumina library preparation adapter sequences.
RNAseq reads were mapped on Svevo cDNA using kallisto (Bray et al., 2016). Basically, index was produced from Svevo cDNA sequence and reads for each sample were mapped on the index Svevo file with default parameters, adding the *--bias* option. The resulted abundance file was analyzed with the R package *sleuth* (Pimentel et al., 2017), adding the experimental design and performing bootstrap analysis. Expression values (transcript per million - tpm) for each gene transcript variant in every sample were calculated. Transcript variant for each gene were merged together and only significant differentially expressed genes (0.05 p-value threshold) were retained for further analysis. The genes within the s*bm2* interval were isolated from Svevo region between 13 Mb and 16 Mb on chromosome 2BS. Differential expressed genes plot was performed using *pheatmap* R package and cluster with Ward.D2 algorithm.

The alignment file was obtained mapping the RNAseq reads on Svevo Ref v1.0 Pseudomolecules with HISAT2 (Kim et al., 2021) aligner. HISAT2 index was built on reference genome and reads were mapped using default parameters with -min-intron-length 40. The bam files were obtained for each sample and used for further analysis.

5.2.8 Svevo RefSeq v1.0 liftover on Svevo Platinum pseudomolecule

The Svevo Platinum pseudomolecule was available from Unibo background material (Maccaferri et al., *unpublished*). The Svevo RefSeq v1.0 transcript annotations were lifted over Svevo Platinum pseudomolecule using GMAP (Wu and Watanabe, 2005) in order to be used for improved gene interval exploration and updated RNAseq analysis.
At first, the transcripts annotated on the Svevo RefSeq v1.0 gff file were transferred on the Svevo Platinum Pseudomolecule, 900 genes had ambiguous annotation (based on different transcript variant on different chromosomes or position) and were manually blasted on the Svevo RefSeq v1.0 cDNA recovering 809 out of 900 genes and generating a new lifted gff transcript annotation file of Svevo Platinum.
Subsequently, this lifted annotation gff was used by HISAT2 to build the Svevo Platinum index and adjust the RNAseq reads mapping on the Svevo Platinum pseudomolecule with shifted gff from Svevo RefSeq v1.0. Finally, reads mapped inside the *sbm2* interval on Svevo Platinum pseudomolecule were used to perform a de novo reference-based transcriptome assembly and annotation using software Stringtie with default parameters. The final transcriptome, based on the Svevo RefSeq v1.0 lifted on Svevo Platinum genome assembly and implemented with de novo annotations at the *sbm2* interval was used to re-analyze RNAseq data using kallisto software and sleuth R package.
Svevo Platinum unmapped reads were further analysed and re-mapped with HISAT2 on SBCMV RNA1 and RNA2 and on vector *Polymixa betae* genomes, with the objective to verify the presence/absence of viral or vector RNAs inside sample roots.
The Svevo RefSeq v1.0 and Svevo Platinum 2B chromosomes were annotated also to detect presence of NLR genes, among the same candidates of resistance to pathogens, using the NLR annotator software (Steuernagel et al., 2020) with default parameters. Genes with partial annotation, annotated as pseudogene or with incomplete NLR domains were discarded, keeping only complete genes.

5.2.9 Comparison between Svevo RefSeq v1.0 and Svevo Platinum

The 2B chromosome on Svevo RefSeq v1.0 was compared to the 2B chromosome on Svevo Platinum in order to check differences between the two assemblies, using mummer 3.23 software (Kurtz et al., 2004) with *Nucmer* function.
Nucmer alignment was carried out for each chromosome using Svevo RefSeq v1.0 as reference and Svevo_Platinum as query, with the *--mum* options to retain only unique matches between reference and query. The output delta file was filtered using the delta-filter command with the following options: -l 10000 to keep aligments longer that 10 kbp between reference and query and -r and -q, which retain only alignment that form the long consistent set for reference and query respectively. First of all, the filtered output was used to generate the plot with the mummerplot command line, which converts the alignment data into a dot plot format exploiting gnuplot. Mummerplot was used with the *--large* option, to set the output size, and *--layout* to orient and order the sequences such

that the largest alignments hit the plot cluster near the main diagonal. The output plot shows the reference sequence (x axis) and the query sequence (y axis) where the bp distance is reported, if the alignment is colinear the dots are displayed in red, whereas inversions are shown in blue.
The delta-filter output file was used also to create a tabular file displaying the summary information about the alignment with the show-coords function. The included options used were *-c* to add the percent coverage in the output, *-l* to add the sequence length in the output, *-T* to convert in tab format, *-d* to include the alignment direction (-1 or 1) and *-r* to order the alignments based on the reference. The output file reports the following columns: start and end of reference sequence alignment, start and end of query sequence alignment, length of the alignment of the reference and for the query, identity percentage of the alignment, length of the reference and of the query, percentage coverage of the reference and query and finally frame orientation for the reference and query.

5.2.10 Statistical analysis on phenotypic data

Analysis of variance (ANOVA) was performed to analyse phenotypic data collected for each genotype, introducing replicates and position of each genotype in the field, interactions between genotypes and replicates (G x E) and environment (E), the analysis was performed with softwares R and Rstudio (Rgroup, 2020) using *car* package. Outliers were removed using the interquartile rules: the interquartile range (IQR) of the data was multiplied by 1.5, outliers were defined as values 1.5 *IQR above the third quartiles and 1.5 * IQR below the third quartile. The position of each genotype in the field was given by XY coordinates where X represents rows and Y columns. BLUEs (Best Linear Unbiased Estimators) were calculated for the phenotypic traits in order to standardize the data, taking in consideration replicates and field positions using the mixed model *lmer* function in *lme4* package (Bates et al., 2015), with the following formula (using genotypes as fixed and other parameters as random):

Symptom_severity ~ Genotype + Replicate + Rows + Columns

Heritability was calculated on phenotypic traits for each replicate using the R package *heritability.*

5.2.11 Comparison of markers order in *sbm2* region among genomes

Markers order between IWB73347 (KUBO 13, position: 9,986,862 bp) and IWA7936 (position: 24,678,860 bp) based on Svevo reference genome was compared with same markers' order (when marker was present) on bread wheat reference genome (Chinese Spring, IWGSC RefSeq V2.1; Zhu et al., 2021) and wild emmer wheat reference genome (Zavitan, WEW_v2.0; Zhu et al., 2019). This physical region comprises IWB10512 (KUBO 9, position on Svevo genome: 24,313,744), the marker flanking *sbm2* along with KUBO 13.
Marker physical position on the durum wheat genome assembly (Maccaferri et al., 2019) was determined by blastn analysis. BLAST version 2.11.0 was run using the flanking marker sequences as queries with an e-value threshold of 10-10 threshold. To solve multiple hits issues due to homoeologues/paralogs matches, for each query, the best hit was defined as the hit with the smallest e-value and the longest High-scoring Segment Pair (HSP) alignment. Hits of markers in the confidence interval were manually cross-checked. Similar blast analyses were applied to determine

the physical position of markers on the Wild Emmer Wheat Zavitan WEWSeq v2.0 (Zhu et al., 2019) and the Wheat Chinese Spring IWGSC RefSeq v2.1 (Zhu et al., 2021) genome assemblies.
The marker position, in some cases, was also checked using Ensembl Plant database (https://plants.ensembl.org/index.html) or graingenes BLAST (https://wheat.pw.usda.gov/blast/) online tools.
Finally, markers in each genome were ordered based on the physical position and a graphical comparison of physical maps was produced using Pretzel (Keeble-Gagnère et al., 2019).

5.2.12 *Sbm2* haplotypes based on Durum Panel

Haploblocks in the region investigated in the previous section were defined for the accessions of the durum panel 1 (DP1) (Maccaferri et al., 2015) using the software Haploview (Barrett et al., 2005). Accessions were grouped based on the different haplotypes at the fifth block of chromosome 2B, the block encompassing the region most associated with *sbm2* based on the GWAS, containing markers KUBO 27, KUBO 29, KUBO 1, KUBO 3, KUBO 41 and KUBO 40. Box plots were produced for each haplotype at block 5, to study the distribution of the accessions based on adjusted visual score data and ELISA traits.

5.2.13 Genotyping of a panel of worldwide durum wheat accessions

Genomic DNA of 291 additional durum wheat accessions (Innovar dururm panel and DP1) was isolated by means of a CTAB DNA extraction protocol for single tubes. DNA samples were then genotyped with KASP markers KUBO 27, KUBO 29 and KUBO 1 following the protocol previously described (Paragraph 5.2.4). Meridiano, Claudio, Svevo, Ciccio and artificial heterozygous samples were used in two replicates as control. The distribution of SBCMV-resistant and susceptible haplotypes in the whole durum panel, for whose accessions the haplotype genotypic data were already available, was studied based on the accessions' passport data (origin, pedigree and year of release) and Unibo background information on SBCMV resistance.

5.2.14 *Sbm2* gene interval exploration

The *sbm2* confidence interval was analyzed in Ensembl plant database using the Biomart tool (Bolser et al., 2016), downloading the genes from *Triticum turgidum* cv Svevo RefSeq v1.0 (Maccaferri et al., 2019). The gene network for the candidate genes was studied obtaining the orthologues on *Triticum aestivum* cv Chinese Spring v1.0 genome, and the gene network was enquired using the knetminer database (Hassani-Pak et al., 2021).

5.3 Results

5.3.1 KASP genotyping

The GWAS analysis and fine mapping performed by Bruschi et al. (*unpublished*) on durum panel and RILs populations, confirmed the *sbm2* QTL presence (Maccaferri et al., 2011b) on durum panel (Maccaferri et al., 2015) for multiple years of phenotyping. Furthermore, SNP markers from Illumina Infinium 90K SNP Chip were converted in KASP markers and tested on the biparental population Meridiano (R) x Claudio (S) and Svevo (R) x Ciccio (S). The tested KASP markers showed a strong association in *sbm2* QTL peak on durum panel and allowed to narrow down the interval on 2B based on informative recombination on the two biparental population (Table 33) (Bruschi et al., *unpublished*).

Table 33: KASP primers converted from the most associated SNPs on Sbm2 QTL. For each KASP marker is reported the genetic and physical position on Svevo genome, the alleles and alleles frequencies and the -logP and R^2 values obtained from GWAS using Visual score data and ELISA on durum panel.

						Visual Score (VS)		ELISA	
Marker	**Position (cM)**	**Position (bp)**	**Strand**	**Allele**	**Allele frequency**	**-Log10(P-Value)**	**R^2 (%)**	**-Log10(P-Value)**	**R^2 (%)**
KUBO 13	8.4	9,986,862	+	A/G	0.45/0.45	5.65	8.3	3.72	5.7
KUBO 27	11.6	13,011,332	+	A/G	0.56/0.39	19.36	35.0	17.76	35.6
wPt-2106	10.9	14,314,449	-	A/T	0.38/0.59	18.48	31.2	15.10	27.8
KUBO 29	11.6	15,643,691	+	G/T	0.57/0.38	18.84	32.4	16.12	30.9
KUBO 1	12.2	15,672,582	-	C/T	0.56/0.26	14.11	34.0	15.49	35.9
KUBO 3	12.3	15,805,908	+	C/T	0.26/0.57	11.99	26.0	15.78	36.4
KUBO 41	12.3	16,184,452	-	C/T	0.68/0.25	6.27	9.3	10.58	19.6
KUBO 40	12.3	16,185,445	+	C/T	0.58/0.25	9.35	19.2	11.75	25.6
KUBO 9	19	24,313,744	+	A/G	0.26/0.66	9.80	16.9	7.17	12.6

The new fine mapped interval was then included between markers KUBO 27 and KUBO 41 narrowing down the QTL interval to 3 Mb, between 13 Mb and 16 Mb on short arm of chromosome 2B (Bruschi et al., *unpublished*).

In order to fine map the *sbm2* interval, a wider chromosomic region, between markers KUBO 9 and KUBO 13 (approximately between 9 and 24 Mb on chromosome 2BS), was explored with KASP developed from Affimetrix SNP chip 420K. KUBO 9 and KUBO 13 determine an interval of approximately 6 cM, SNP markers from 420K Affimetrix SNP Chip within the interval were converted to KASP primers at first with Polymarker tool (Ramirez-Gonzalez et al., 2015) 11 markers in total showed to be polymorphic, tested with SBCMV resistant or susceptible durum lines and artificial

heterozygotes. KASP markers with separated allelic varieties were validated as polymorphic, their position in the recombination bin was confirmed with M x C RILs population genotyping. On the other hand, KASP markers that showed dominance or that were not polymorphic were re-designed by multiple alignment sequences of Svevo, Chinese Spring and Zavitan. KUBO 63, KUBO 64, KUBO 65, KUBO 68, KUBO 72, KUBO 73, KUBO 74, KUBO 75, KUBO 76, KUBO 80, KUBO 81 and KUBO 82 were designed inside the confidence interval KUBO 9 / KUBO 13 and proved to be polymorphic from analysis on resistant and susceptible durum lines. 11 markers were converted from 420K SNP Chip Affimetrix and KUBO 80 from SNP90K Chip consensus map (Maccaferri et al., 2015).
These markers, together with KUBO 29 (Bruschi et al., *unpublished*) and DArT marker wpt-2106 were used to genotype MxMC 600 RILs population in order to find new recombination events.
Also in this case, the KUBO markers were tested with durum wheat resistant and susceptible varieties, also an artificial heterozygote was analysed combining genotypes with opposite haplotype as shown in the amplification plots (Figure 44). KUBO 73, KUBO 76, KUBO 80 and KUBO 81 showed a sufficient discrimination capacity between resistant and susceptible cultivars but not so evident on the artificial heterozygote, so amplification plots for these markers are not included in this report. Furthermore, by a BLAST analysis on Svevo genome latest version, KUBO 73 and KUBO 80 have strong hits also on A genome (chr4A and chr2A respectively) that could be one cause of a not completely haplotype discrimination. Therefore, the markers showed in the amplification plots could be ascribed as the most consistent for *sbm2* analysis.

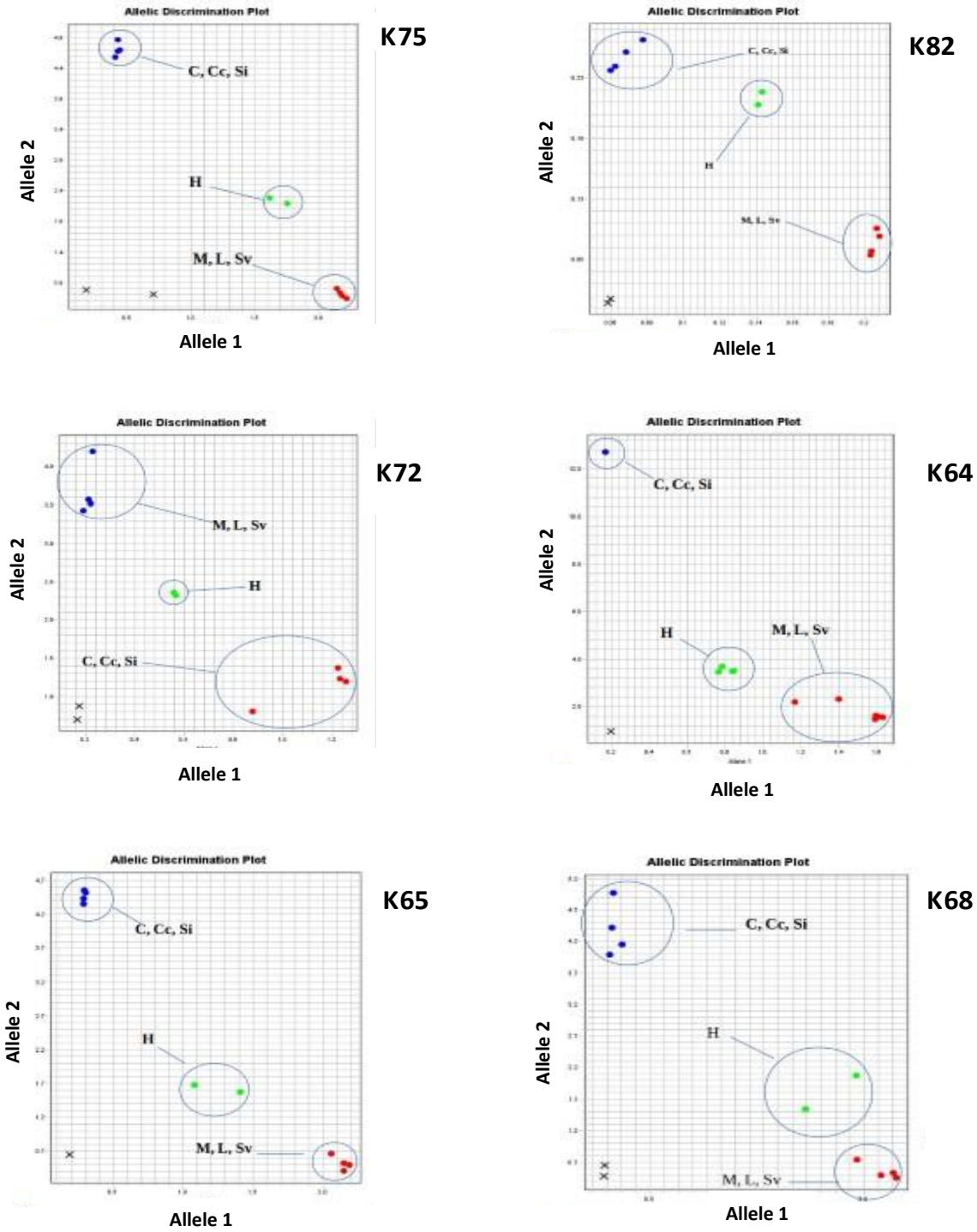


Figure 44: The amplification plots show good discrimination capacity among different markers between resistant/susceptible genotypes and artificial heterozygotes. Varieties tested are the following: C=Claudio, Cc=Ciccio, Si=Simeto (susceptible varieties); M=Meridiano, L=Levante, Sv=Svevo (resistant varieties).

The Affimetrix KASP markers genotyping on MxMC BCF_4 RILs resulted in 13 recombinant lines between KUBO 74 and KUBO 9 (Figure 45).

MxMC	KUBO74	KUBO27	KUBO64	KUBO75	KUBO65	wPt-2106-2B	KUBO63	KUBO76	KUBO80	KUBO72	KUBO68	KUBO29	KUBO81	KUBO82	KUBO73	KUBO3	KUBO9
1096-3	NA	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B
1089-1	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A
1033-3	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B
1033-2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B
1011-4	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B
1002-1	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
920-3	NA	B	B	B	B	B	B	B	NA	B	B	B	B	B	B	B	A
171-4	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
57-1	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A
62-6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H
75-5	B	B	B	B	NA	B	NA	B	B	B	B	B	B	B	B	B	A
2252-4	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A
1365-1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2252-3	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A
219-5	NA	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
265-5	NA	B	B	NA	NA	B	B	B	B	B	B	B	B	B	B	B	B
1290-2	B	A	A	NA	A	A	A	A	A	A	A	A	A	NA	A	A	A
meridiano	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
meridiano	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
claudio	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
claudio	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

Figure 45: confidence interval of MxMC population showing genotyping results of KASP markers assays, developed from Illumina 90K SNP Chip and Affimetrix 420K SNP array. "A" stands for resistant allele, "B" stands for susceptible allele, "H" stands for heterozygotes and missing data are reported with "NA".

In order to study the different haplotypes, a comparison between the Illumina SNP 90K markers order in 3 reference genomes (physical maps) of *Triticum turgidum* (Svevo), *Triticum dicoccoides* WEW v2.0 (Zavitan) and *Triticum aestivum* IWGSC RefSeq v1.0 (Chinese spring) was performed using Pretzel database. As reported in Figure 46, the marker order was maintained between three genomes with only few rearrangements.

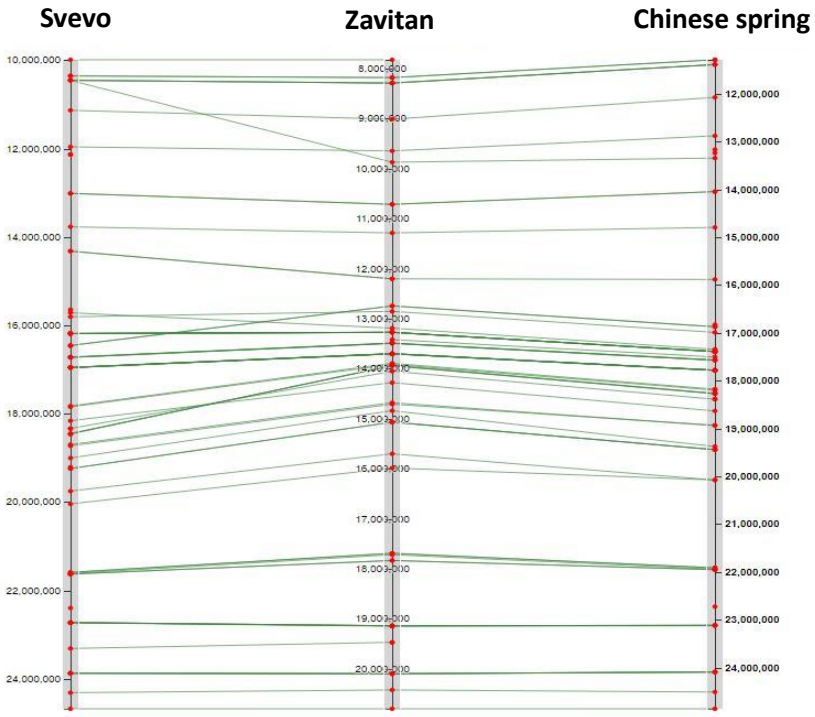


Figure 46: QSbm.ubo-2BS region, Chr. 2BS, Mb 9-24. 90K SNP markers order comparison between three reference genomes: Triticum turgidum (Svevo), Triticum dicoccoides WEW v2.0 (Zavitan) and Triticum aestivum IWGSC RefSeq v1.0 (Chinese spring).

5.3.2 Phenotypic analysis

The MxMC BCF_4 population was evaluated in three different years: 2018 (background data already available before the PhD beginning), 2019 and 2021. During these three years symptom severity (SS) was evaluated in different dates, with the only exception of 2019 where the ELISA test was performed as well. BLUES values were calculated for all the years and descriptive statistics was computed, together with heritability.
SS score was evaluated during spring 2018 in six different dates from March to April. Approximately 138 MxMC RILs genotypes were evaluated in the Cadriano (Unibo) field trial, sown in two replicates, out of initial 600 lines. In fact, as a consequence of the KUBO 9 and KUBO 13 fine mapping (Bruschi et al., *unpublished*), only a sub selection of recombinant lines was tested in the field.
During autumn 2018, after KASP genotyping, further selected MxMC recombinants and additional genotypes (to confirm 2017 phenotypic values) were sowed in SBCMV field in three replicates, a total of 56 genotype for each replicate. SS was scored in four different dates from March to April 2019 and BLUES values were calculated using genotype coordinates in the field to adjust for field effects. In addition, ELISA test was performed on all parcels, sampling 12 random leaves for every genotype in one single date (April 2019). ELISA values were then normalized using the positive control sample with the highest value. The experiment was repeated in 2021, with the same number of varieties used in 2019, in two replicates with randomized block design. A summary of descriptive statistics of the three years of trial is reported in Table 34.

Table 34: descriptive statistic for ELISA test for MxMC population, year 2018, 2019 and 2021. BLUES values were calculated for each genotype across the different traits. The heritability values for each year trial were included.

	MxMC VS 2018	**MxMC VS 2019**	**MXMC ELISA 2019**	**MXMC VS 2021**
Average	**3.64**	1.85	0.44	2.68
Standard error	0.037	0.061	0.059	0.07
Standard deviation	0.41	0.46	0.44	0.63
Skewness	-0.34	0.045	0.64	0.31
Min	2.54	1.007	-0.001	1.50
Max	4.42	2.82	1.28	4.13
CV	0.11	0.25	1.013	0.23
h^2	0.49	0.28	0.7	0.89

As regards the heritability values, 2019 season was lower than expected. A probable cause was the low disease presence during the season and the concurrent presence of cold stress that hamper the correct visual evaluation of disease symptoms. However, on the same year, the ELISA test was performed, and it gave a better outcome regarding the disease presence and heritability values (0.7). Therefore, the ELISA values were considered the most reliable for 2019 in comparison to the visual score. The highest disease pressure was detected during the 2021 growing season, with the

highest heritability value (0.89) obtained from the analysis of visual score data performed on 4 evaluation dates.

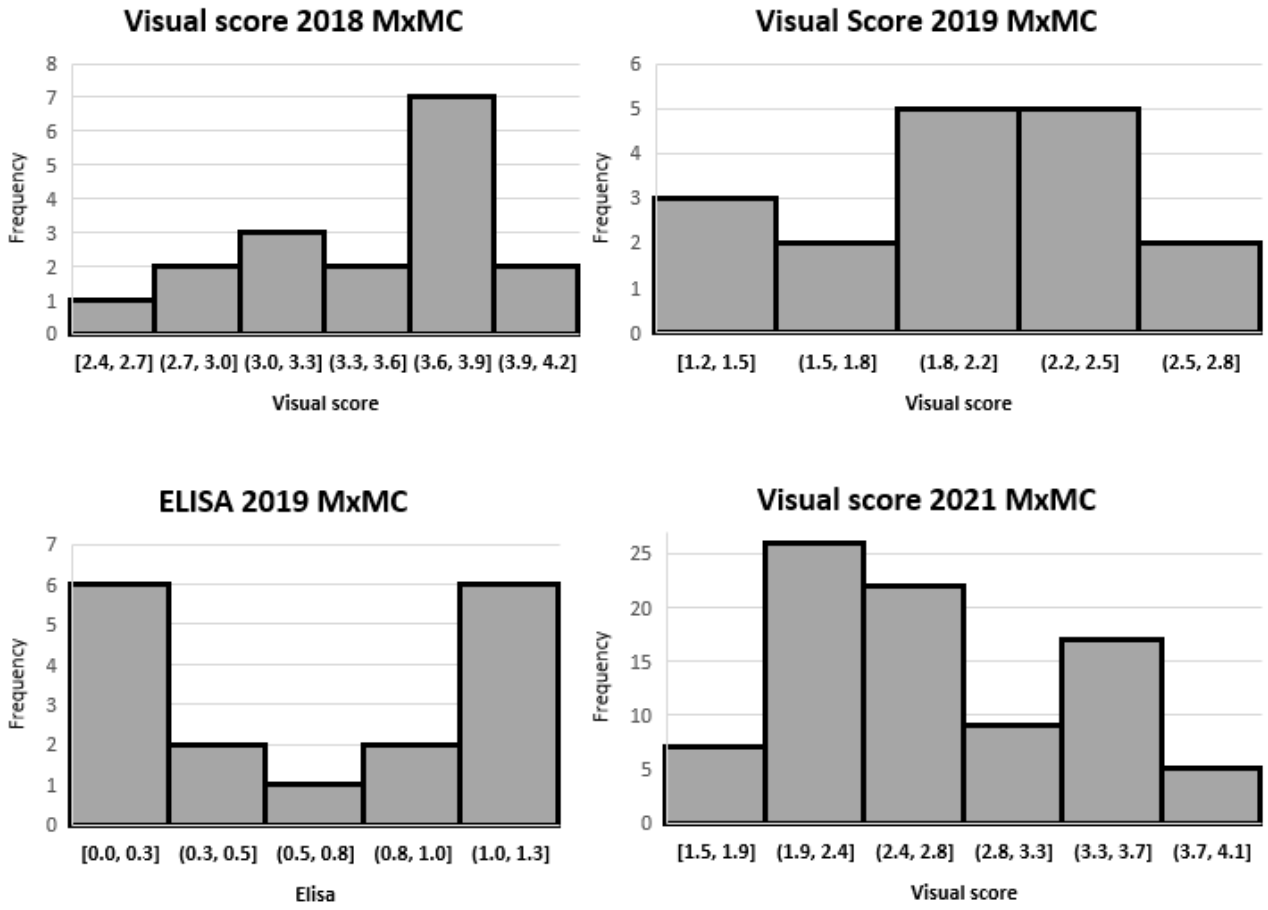


Figure 47: phentoypic distribution of Visual score and Elisa values for 2018, 2019, 2021 of MxMC RILs selected population.

The phenotypic distribution of the multi-year SS traits could be ascribed to a bimodal distribution (Figure 47). This leads to the fact that the resistance to the SBCMV in the MxMC population has a strong genetic component which discriminates precisely between resistant and susceptible genotypes, with high segregation in the different MxMC generations.

5.3.3 Fine mapping of the *sbm2* QTL

Phenotypic data were merged on genotypic data from MxMC BCF_4 population, genotyped with 11 *KASP* markers developed from Affimetric 420K SNP Chip array. As detected by comparing phenotypic values from different years (2018 and 2019), the ELISA seemed to better discriminate resistant and susceptible varieties as proved also by the bimodal distribution of the phenotype. Based on the phenotypic and genotypic results (excluding VS from 2019 because of the low heritability), the confidence interval could be narrowed down between markers KUBO 63/KUBO 76 and markers KUBO 29/KUBO 81 (Figure 48).

Fine mapped interval on MXMC population RIL

MXMC110 ID	KUBO74	KUBO27	KUBO64	KUBO75	KUBO65	wPt-2106-2B	KUBO63	KUBO76	KUBO80	KUBO68	KUBO72	KUBO29	KUBO73	KUBO81	KUBO82	KUBO3	KUBO9	BLUE_VS_2018	BLUE_ELISA_2019	BLUES_VS_2021
1002-1	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	3.85	0.82	3.40
1033-2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	3.50	0.38	2.67
1033-3	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	3.08	0.18	2.23
1089-1	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	3.67	1.28	3.52
1096-3	NA	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	3.72	1.11	3.45
1011-4	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	3.24	0.14	2.20
1290-2	B	A	A	NA	A	A	A	A	A	A	A	A	A	A	NA	A	A	2.87	0.15	2.35
62-6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	3.04	0.01	1.85
1365-1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	2.70	0.26	2.27
171-4	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	4.05	0.94	3.21
219-5	NA	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	4.15	1.19	3.63
2252-3	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	3.43	0.23	2.01
2252-4	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	2.38	0.44	2.19
265-5	NA	B	B	NA	NA	B	B	B	B	B	B	B	B	B	B	B	B	3.71	1.27	2.75
57-1	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	3.80	0.73	NA
75-5	B	B	B	B	NA	B	NA	B	B	B	B	B	B	B	B	B	A	3.83	1.06	3.00
920-3	NA	B	B	B	B	B	B	B	NA	B	B	B	B	B	B	B	A	3.69	1.28	3.42

Figure 48: Fine mapping on MxMC recombinants using the KASP markers developed inside the Sbm2 interval (13-16 Mb). "A" stands for resistant allele, "B" for susceptible allele, "H" for heterozygotes and missing data as "NA". The phenotypic values reported come from Visual score analysis of 2018 and 2021, and ELISA od 2019. The fine mapped interval was detected between markers KUBO63/76 and KUBO29/81, based on the MxMC critical recombinants: 2252-3, 2252-4, 1011-4, 1096-3, 1033-2 and 1033-3

As a consequence, KASP markers were aligned (*blastn*, NCBI blast) against Svevo RefSeq v1.0 Pseudomolecule, hits were filtered for more than 70% identity and top 5 hits were retained, choosing the best hit based on e-value, highest query coverage and identity.
Based in these results, the restricted interval on chromosome 2B was detected between 14.7 Mb and 15.7 Mb, fine mapping the interval to 1.1 Mb (Table 35).

Table 35: KASP markers position on Svevo RefSeq v1.0 pseudomolecule. The fine mapped interval is highlighted in blue (14.7/15.7 Mb), and it was narrowed down to 1.1 Mb in comparison to the previous interval of 3 Mb.

KUBO name	blast on Svevo RefSeq v1.0
KUBO 74	13390403
KUBO 64	13385437
KUBO 75	13391443
KUBO 65	13762796
KUBO 63	14700984
KUBO 76	15592223
KUBO 72	15631168
KUBO 68	15629783
KUBO 29	15643691
KUBO 81	15709012
KUBO 82	15712076

The genetic interval on Svevo chr2B chromosome between markers KUBO 13 and KUBO 41 was explored using the Ensembl Plants database. The narrowed down region between markers KUBO 63 and KUBO 81 includes 30 genes among high and low confidence genes (Table 36). Based on the protein functions ascribed for each gene, the following pathways could be candidate for resistance to SBCMV: cytochrome p450, protein kinases, defensins and NBS-LRR resistant proteins. NBS-LRR protein coding genes can be ascribed as the principal candidates for resistance to pathogens, based on the Svevo interval, genes with this function aggregate together in the annotation, being major candidates for the resistance, namely: *TRITD2Bv1G007220*, *TRITD2Bv1G007240* and *TRITD2Bv1G007250*.

Table 36: Sbm2 fine mapped interval between KUBO27 and KUBO 41 (Bruschi et al., manuscript in preparation). The genes with relative functions are reported, together with KASP marker position. The candidate genes for resistance reaction are represented in bold.

Genes_marker	start	end	Protein_function
KUBO13	9986862	9986962	NA
KUBO80	10455079	10455160	NA
IWB11421	13011332	13011409	NA
KUBO27	13011332	13011409	
TRITD2Bv1G006510	13011439	13015688	E3 ubiquitin ligase BIG BROTHER-related protein
TRITD2Bv1G006520	13019945	13024532	Formamidopyrimidine-DNA glycosylase
TRITD2Bv1G006530	13088442	13089065	Retrovirus-related Pol polyprotein from transposon TNT 1-94
TRITD2Bv1G006540	13128745	13129163	ATP synthase subunit b 2
TRITD2Bv1G006550	13144764	13145612	Terpene synthase
TRITD2Bv1G006560	13188067	13188369	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G006570	13264020	13264445	Gag-Pol polyprotein
TRITD2Bv1G006580	13291123	13291791	CDT1-like protein a, chloroplastic
TRITD2Bv1G006590	13305904	13308322	Myrcene synthase, chloroplastic
TRITD2Bv1G006600	13382910	13385424	Myrcene synthase, chloroplastic
KUBO64	13385437	13385507	NA
TRITD2Bv1G006610	13389697	13392529	Cytochrome P450
KUBO74	13390403	13390473	NA
TRITD2Bv1G006620	13390645	13391031	ARM repeat superfamily protein
KUBO75	13391443	13391513	NA
TRITD2Bv1G006630	13409178	13409840	RNA-directed DNA polymerase (reverse transcriptase)-related family protein
TRITD2Bv1G006640	13414501	13416024	Myrcene synthase, chloroplastic
TRITD2Bv1G006650	13417163	13417723	Myrcene synthase, chloroplastic
TRITD2Bv1G006660	13435526	13437197	Myrcene synthase, chloroplastic
TRITD2Bv1G006670	13451142	13453526	Myrcene synthase, chloroplastic
TRITD2Bv1G006680	13488438	13490857	Cytochrome P450 family protein, expressed
TRITD2Bv1G006690	13537757	13540226	Myrcene synthase, chloroplastic
TRITD2Bv1G006700	13761882	13768700	Cytochrome P450
KUBO65	13762796	13762867	NA
TRITD2Bv1G006710	13781131	13781603	Myrcene synthase, chloroplastic
TRITD2Bv1G006720	13783271	13783642	Retrovirus-related Pol polyprotein from transposon TNT 1-94
TRITD2Bv1G006730	13805895	13808737	ATP-dependent Clp protease adapter protein ClpS
TRITD2Bv1G006740	13835945	13836244	lysine ketoglutarate reductase trans-splicing-like protein (DUF707)
TRITD2Bv1G006750	13861376	13861726	Retrovirus-related Pol polyprotein from transposon TNT 1-94
TRITD2Bv1G006760	13862402	13863763	Retrovirus-related Pol polyprotein from transposon TNT 1-94
TRITD2Bv1G006770	13876446	13877171	Retrotransposon protein, putative, Ty3-gypsy subclass
TRITD2Bv1G006780	13912711	13913079	PQ-loop repeat family protein / transmembrane family protein
TRITD2Bv1G006790	13913841	13916411	Cytochrome P450

TRITD2Bv1G006800	13937938	13938597	Myosin regulatory light chain 2, skeletal muscle isoform
TRITD2Bv1G006810	13956898	13957218	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G006820	14052568	14054346	Cytochrome P450
TRITD2Bv1G006830	14082288	14083965	Cytochrome P450
TRITD2Bv1G006840	14148344	14148649	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G006850	14199183	14199527	Transposon protein, putative, Mutator sub-class, expressed
TRITD2Bv1G006860	14222013	14222336	GMP synthase [glutamine-hydrolyzing]
TRITD2Bv1G006870	14263143	14263484	Retrotransposon protein, putative, unclassified
TRITD2Bv1G006880	14275944	14276273	NAD(P)-binding Rossmann-fold superfamily protein
TRITD2Bv1G006890	14318626	14320342	Cytochrome P450
TRITD2Bv1G006900	14323721	14324242	Protein translocase subunit SecA 1
TRITD2Bv1G006910	14324678	14325224	zinc/iron-chelating domain protein
TRITD2Bv1G006920	14345071	14345643	Cytochrome P450 family protein, expressed
TRITD2Bv1G006930	14345729	14346193	Cytochrome P450
TRITD2Bv1G006940	14346917	14347279	Cytochrome P450
TRITD2Bv1G006950	14350222	14350614	Cytochrome P450
TRITD2Bv1G006960	14350624	14351470	Cytochrome P450
TRITD2Bv1G006970	14360518	14361090	Cytochrome P450
TRITD2Bv1G006980	14362137	14362739	Cytochrome P450
TRITD2Bv1G006990	14409959	14410585	Cytochrome P450
TRITD2Bv1G007000	14420662	14420997	Copia-like polyprotein
TRITD2Bv1G007010	14429622	14430922	Protein kinase superfamily protein
TRITD2Bv1G007020	14451870	14470317	Cytochrome P450 family protein, expressed
TRITD2Bv1G007030	14457084	14457850	Cytochrome P450
TRITD2Bv1G007040	14484180	14484590	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G007050	14501192	14501656	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G007060	14512481	14514197	Cytochrome P450
TRITD2Bv1G007070	14700417	14702116	Myrcene synthase, chloroplastic
KUBO63	14700984	14700914	
TRITD2Bv1G007080	14702343	14702892	Myrcene synthase, chloroplastic
TRITD2Bv1G007090	14724839	14725354	Pol polyprotein
TRITD2Bv1G007100	14728583	14735933	TSL-kinase interacting protein 1
TRITD2Bv1G007110	14777445	14778155	Retrovirus-related Pol polyprotein from transposon TNT 1-94
TRITD2Bv1G007120	**14794208**	**14795918**	**Cytochrome P450**
TRITD2Bv1G007130	14824480	14824824	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase
TRITD2Bv1G007140	14824863	14825294	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
TRITD2Bv1G007150	**14851144**	**14852099**	**Serpin-like protein**
TRITD2Bv1G007160	**14852299**	**14852637**	**Serpin-like protein**
TRITD2Bv1G007170	**14857557**	**14859608**	**Receptor-like kinase**
TRITD2Bv1G007180	**14868843**	**14869181**	**Serpin-like protein**
TRITD2Bv1G007190	**14869381**	**14870336**	**Serpin-like protein**
TRITD2Bv1G007200	14925817	14926725	Omega-6 fatty acid desaturase, endoplasmic reticulum isozyme 2

TRITD2Bv1G007210	14930339	14930749	Retrovirus-related Pol polyprotein from transposon TNT 1-94
TRITD2Bv1G007220	**14944520**	**14948411**	**NBS-LRR-like resistance protein**
TRITD2Bv1G007230	**15010013**	**15022097**	**receptor kinase 1**
TRITD2Bv1G007240	**15053487**	**15080119**	**NBS-LRR-like resistance protein**
TRITD2Bv1G007250	**15081109**	**15087938**	**NBS-LRR disease resistance protein-like**
TRITD2Bv1G007260	**15092451**	**15096269**	**receptor kinase 1**
TRITD2Bv1G007270	15113911	15114856	RING/FYVE/PHD zinc finger superfamily protein
TRITD2Bv1G007280	15122880	15123212	(DL)-glycerol-3-phosphatase 2
TRITD2Bv1G007290	15152588	15153496	Integrase-type DNA-binding superfamily protein
TRITD2Bv1G007300	15154563	15155545	Jasmonate-induced protein
TRITD2Bv1G007310	15181521	15181955	piezo-type mechanosensitive ion channel component
TRITD2Bv1G007320	15193086	15193545	ABC-2 type transporter family protein
TRITD2Bv1G007330	15271707	15272099	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G007340	15328195	15328965	Retrotransposon protein, putative, unclassified
TRITD2Bv1G007350	15470221	15470562	Retrotransposon protein, putative, Ty3-gypsy subclass
TRITD2Bv1G007360	15541650	15541994	Transposon protein, putative, CACTA, En/Spm sub-class
TRITD2Bv1G007370	15585810	15587738	Receptor-like protein kinase
TRITD2Bv1G007380	15589495	15590559	Glycosyltransferase
KUBO76	15592223	15592293	
KUBO68	15629783	15629853	
TRITD2Bv1G007390	**15630524**	**15806566**	**receptor kinase 1**
TRITD2Bv1G007400	**15630833**	**15632470**	**Receptor-like protein kinase**
KUBO72	15631168	15631238	
KUBO29	15643691	15643791	
TRITD2Bv1G007410	**15644845**	**15645322**	**Defensin**
TRITD2Bv1G007420	**15662923**	**15663300**	**Defensin**
TRITD2Bv1G007430	**15684330**	**15686489**	**Receptor-like protein kinase**
TRITD2Bv1G007440	**15707161**	**15709249**	**GDSL esterase/lipase**
KUBO81	15709012	15709082	
TRITD2Bv1G007450	**15711304**	**15713726**	**Carboxyl methyltransferase**
KUBO82	15712076	15712146	
KUBO73	15720532	15720602	NA
TRITD2Bv1G007460	15721167	15723131	Protein kinase, putative
TRITD2Bv1G007470	15784509	15786270	Glutamyl-tRNA (Gln) amidotransferase subunit A
KUBO3	15805958		
TRITD2Bv1G007480	**15838562**	**15841457**	**Receptor-like protein kinase**
TRITD2Bv1G007490	15846893	15847763	Glutamyl-tRNA (Gln) amidotransferase subunit A
TRITD2Bv1G007520	15891737	15893627	Protein MID1-COMPLEMENTING ACTIVITY 1 G
TRITD2Bv1G007540	15905129	15905756	Glutamyl-tRNA (Gln) amidotransferase subunit A
TRITD2Bv1G007550	**15923669**	**15925153**	**Cytochrome P450**
TRITD2Bv1G007560	**15968115**	**15975371**	**Cytochrome P450**
TRITD2Bv1G007570	15976977	15985199	Glutamyl-tRNA (Gln) amidotransferase subunit A
KUBO41	16184388		

In order to further check the gene annotation, the *sbm2* interval from KUBO 27 to KUBO 41 identified by Bruschi et al., (fine mapped interval between 13 Mb and 16 Mb) (Bruschi et al., *unpublished*) was analysed to confirm the presence of NBS-LRR protein coding genes using NLR-annotator tool (Steuernagel et al., 2020).
As reported in Table 37, the NLR protein coding genes annotated on the genome inside the region are confirmed by the NLR annotator tool. The tools detected complete domains as regards the *TRITD2Bv1G007220* and *TRITD2Bv1G007240* only proteins with complete functional domains and p-loops. On the other hand, the *TRITD2Bv1G007250* was appointed as partial NBS due to the absence of p-loop domain in protein structure.
In Table 37, the gene *TRITD2Bv1G006350* is outside the upstream flanking marker KUBO 27 and *TRITD2Bv1G007810* complete gene is outside the downstream flanking marker KUBO 41. As regards this gene, partial NBS-LRR domains are annotated inside the *sbm2* interval, but they are classified as pseudogenes.

Table 37: NLR identified by NLR annotator tool in the Sbm2 confidence interval between KUBO27 and KUBO3 (13-16 Mb) on Svevo Platinum genome. The protein domain is specified as complete, partial or pseudogene.

Genes	domain status	start	end
TRITD2Bv1G006350	**partial**	**12727616**	**12729104**
TRITD2Bv1G007220	**complete**	**14945443**	**14948372**
TRITD2Bv1G007240	**complete**	**15011339**	**15015672**
TRITD2Bv1G007250	**partial**	**15067136**	**15068171**
TRITD2Bv1G007810	**complete**	**15081147**	**15086894**
TRITD2Bv1G007810	**complete (pseudogene)**	**15735966**	**15743539**
TRITD2Bv1G007810	**partial (pseudogene)**	**15773286**	**15776118**

5.3.4 Comparison with Svevo Platinum pseudomolecule

In order to check the correct assembled sequence of Svevo RefSeq v1.0 inside the *sbm2* region, an alignment comparison was performed with the update version of Svevo pseudomolecule assembly available in Unibo (Maccaferri et al., unpublished), namely Svevo Platinum. First of all, the chr2B of Svevo RefSe1 v1.0 was compared to Svevo Platinum using mummer 3.23 package with nucmer function.
The chromosome comparison shows that there are different inversion or duplications especially in the telomeric region of the chromosome, pointing out orientation errors between Svevo RefSeq v1.0 in comparison to Svevo Platinum (Figure 49). Looking in details at the *sbm2* region, it is visible that there is a strong 500 kbp inversion between 14.3 Mb and 14.8 Mb in Svevo RefSeq v1.0 genome which directly involves the fine mapping interval with the flanking marker KUBO 63, at 14,700,984 bp. This leads to the inversion of the region and inclusion of different genes, involved principally in cytochrome p450 pathway, that were previously excluded from the fine mapped interval.

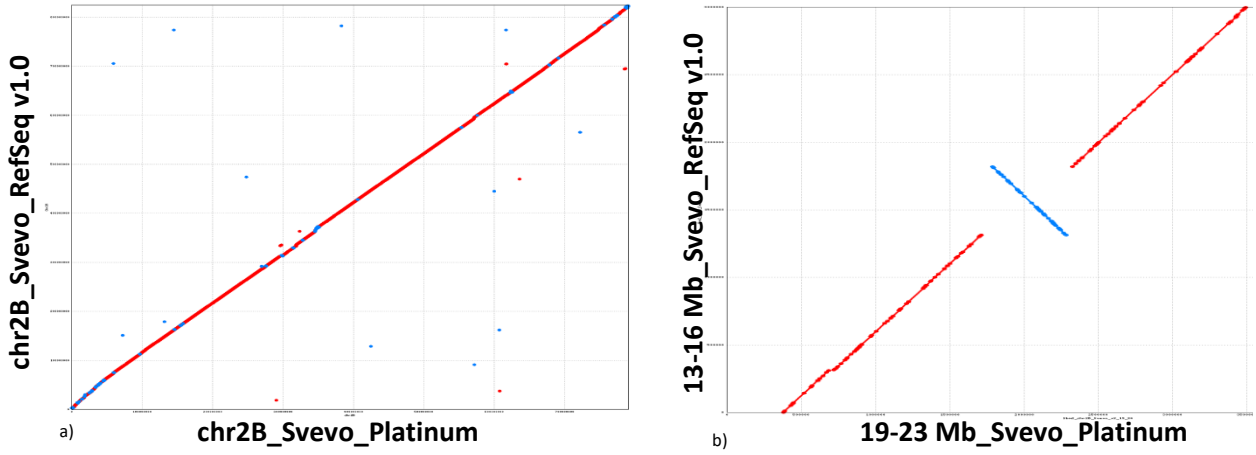


Figure 49: mummer output from the comparison between chr2B Svevo RefSeq v1.0 (query) and chr2B Svevo Platinum (a) and between sbm2 fine mapped regions (b), 13-16 Mb for Svevo RefSeq v1.0 and 19-23 Mb for Svevo Platinum. The red dots show linearity between the two sequences, the blue dots show inversions between query and reference sequence

The *sbm2* corresponding interval on Svevo Platinum is between 19.3 Mb and 22.7 Mb on chromosome 2B, with 300 kbp added to the Svevo Platinum interval that could be a consequence of the Svevo RefSeq v1.0 inversion.
In order to confirm that the wrong oriented interval was corresponding to Svevo RefSeq v1.0, the 2B chromosome of Svevo Platinum was aligned to *Triticum aestivum* cv. Chinese Spring (CS) chromosome 2B, showing the presence of small duplications, but the inversion at the proximal telomeric part of the chromosome was not present (Figure 50).

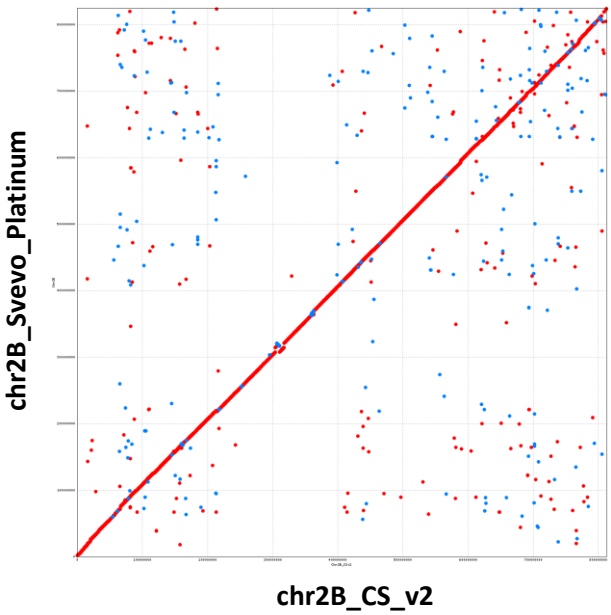


Figure 50: mummer output between chr2B of Svevo Platinum and chr2B of CS RefSeq v2.0. The red dots show linearity between the two sequences, the blue dots show inversions between query and reference sequence.

To confirm the presence of the candidate NLR protein coding gene and to further check the Svevo RefSeq v1.0 annotation, NLR annotator was used in the *sbm2* corresponding region of Svevo Platinum.
The previously annotated NLRs inside the *sbm2* interval of Svevo RefSeq v1.0 were consistent in the Svevo Platinum version, with the only exception of *TRITD2Bv1G007230* gene, previously annotated as low confidence gene with a different function in Svevo RefSeq v1.0 (Table 38).

Table 38: NLR identified by NLR annotator tool in the Sbm2 confidence interval between KUBO27 and KUBO3 (19-24 Mb) on Svevo Platinum genome. The protein domain is specified as complete, partial or pseudogene.

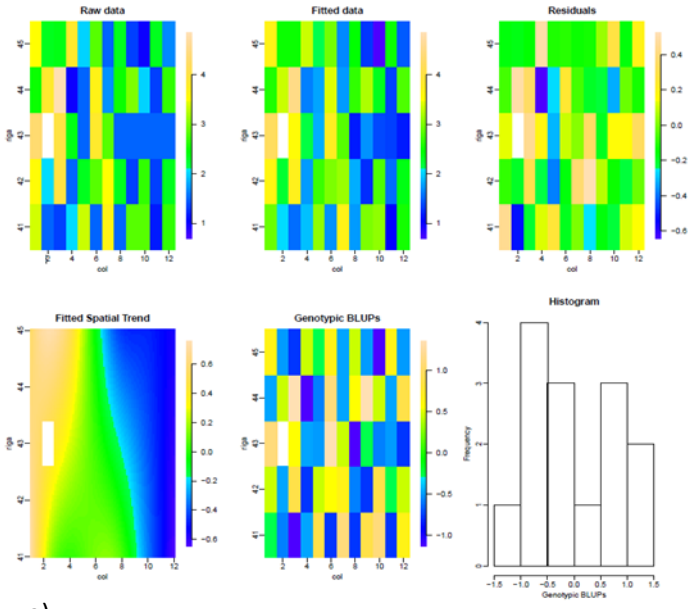
Genes	domain status	start	end
TRITD2Bv1G006350	**partial**	**19,110,739**	**19,112,227**
TRITD2Bv1G007220	**complete**	**21,450,493**	**21,453,422**
TRITD2Bv1G007230	**complete**	**21,514,513**	**21,518,846**
TRITD2Bv1G007240	**partial**	**21,569,796**	**21,570,831**
TRITD2Bv1G007250	**complete**	**21,583,720**	**21,589,467**
TRITD2Bv1G007810	**complete (pseudogene)**	**22,234,773**	**22,242,346**
TRITD2Bv1G007810	**partial (pseudogene)**	**22,261,050**	**22,261,755**
TRITD2Bv1G007810	**partial (pseudogene)**	**22,272,093**	**22,274,925**
TRITD2Bv1G007810	**partial (pseudogene)**	**22,272,093**	**22,274,925**

To conclude, differences between the two genome versions seem to be ascribed only to the 500 kbp inversions between 14.3 Mb and 14.8 Mb which includes genes that were previously excluded. However, the principal candidate NLR protein coding genes are consistent between the two version with the only exception of a gene addition in Svevo Platinum genome version.

5.3.5 Phenotypic and RNAseq analysis

The RNAseq experiment was carried out with root samples collected from the SBCMV field in 3 different dates: 04-02-21, 18-02-21 and 04-03-21. These data were chosen as the progress of the infection was evaluated by visual score and ELISA. Data were analysed and corrected for spatial distribution using Spats R package (Figure 51 a and b).

Visual score (VS) assessment on transcriptomic varieties



a)

Transcriptomics_VS	
Average	2.38
Standard error	0.21
Standard deviation	0.80
Variance	0.63
Kurtosis	-1.36
Skewness	0.36
Min	1.26
Max	3.72
CV	0.33
h^2	0.95

b)

Figure 51: a) Spatial and phenotipic distribution of BLUEs visual score symptoms for durum transcriptomics pane across different dates os SS evaluationl. b) Descriptive statistics of durum transcriptomics panel for visual score BLUEs across different SS evaluation dates.

The infection gradient was visible in the field trial, showing more infection at the right end of the field. However, the analysis corrected the gradient when generating BLUEs, showing the expected bimodal distribution of the phenotype as the trial is composed by resistant and susceptible varieties. The heritability of the trait was about 0.95.

In order to confirm the phenotypic data, the ELISA was performed on the three dates before sampling plants for RNAseq. The phenotypic distribution for the three-sampling point of the three replicates was bimodal, strengthening the strong segregation of the trait between the parental lines (Figure 52). The heritability values for the ELISA traits ranged between 0.8 and 0.93 across different dates.

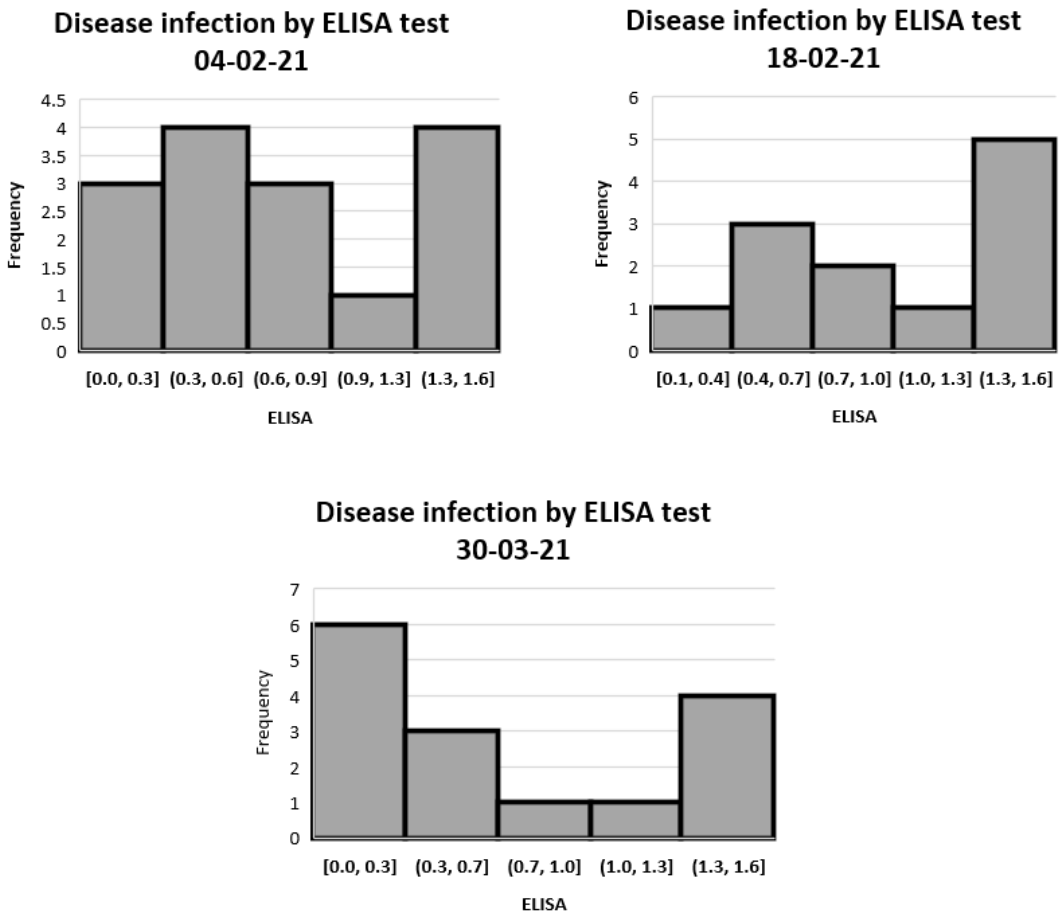


Figure 52: ELISA phenotypic distribution of BLUEs ELISA values across the different dates for the transcriptomic pane.

The RNA experiment was performed on samples of three biological replicates for each sampling date of the following varieties: Svevo (R), Ciccio (S), resistant Bulk (RB: Levante, Neodur, Meridiano) and susceptible bulk (SB: Simeto, Claudio, Altar-84). RNAseq was performed on roots as it is the first tissue that is in contact with the SBCMV and the *Polymixa graminis* vector, so it is the principal candidate tissue where resistance reaction could occur. RNA was extracted and RNAseq was performed on Svevo, Ciccio, RB and SB for 04-02-2121, 18-02-2021 and 04-03-2021. RNAseq output data resulted in between 40 and 80 million of reads pairs sequenced for each sample.
The reads were aligned to the Svevo RefSeq v1.0 annotated cDNAs using *kallisto* and *sleuth* for statistical analysis, merging transcript for each gene and extracting normalized gene expression values (transcript per million, tpm) for each gene in each analysed sample.
The Svevo RefSeq v1.0 was explored in the *sbm2* interval between 13 Mb and 16 Mb on short arm of chromosome 2B. 25 genes inside the interval showed differential expression between resistant and susceptible samples. As shown in Figure 53, resistant and susceptible samples cluster separately and differentially expressed gene clusters are visible: in order, from *TRITD2Bv1G007400* to *TRITD2Bv1G007570* genes were overexpressed in susceptible cultivars, similarly from *TRITD2Bv1G007480* to *TRITD2Bv1G007320* genes were overexpressed in resistant cultivars. Furthermore, three genes seemed to be strongly overexpressed in resistant cultivars in comparison to susceptible ones, namely: *TRITD2Bv1G007390* (protein kinases), *TRITD2Bv1G007240* (NBS-LRR) and *TRITD2Bv1G007260* (protein kinases).

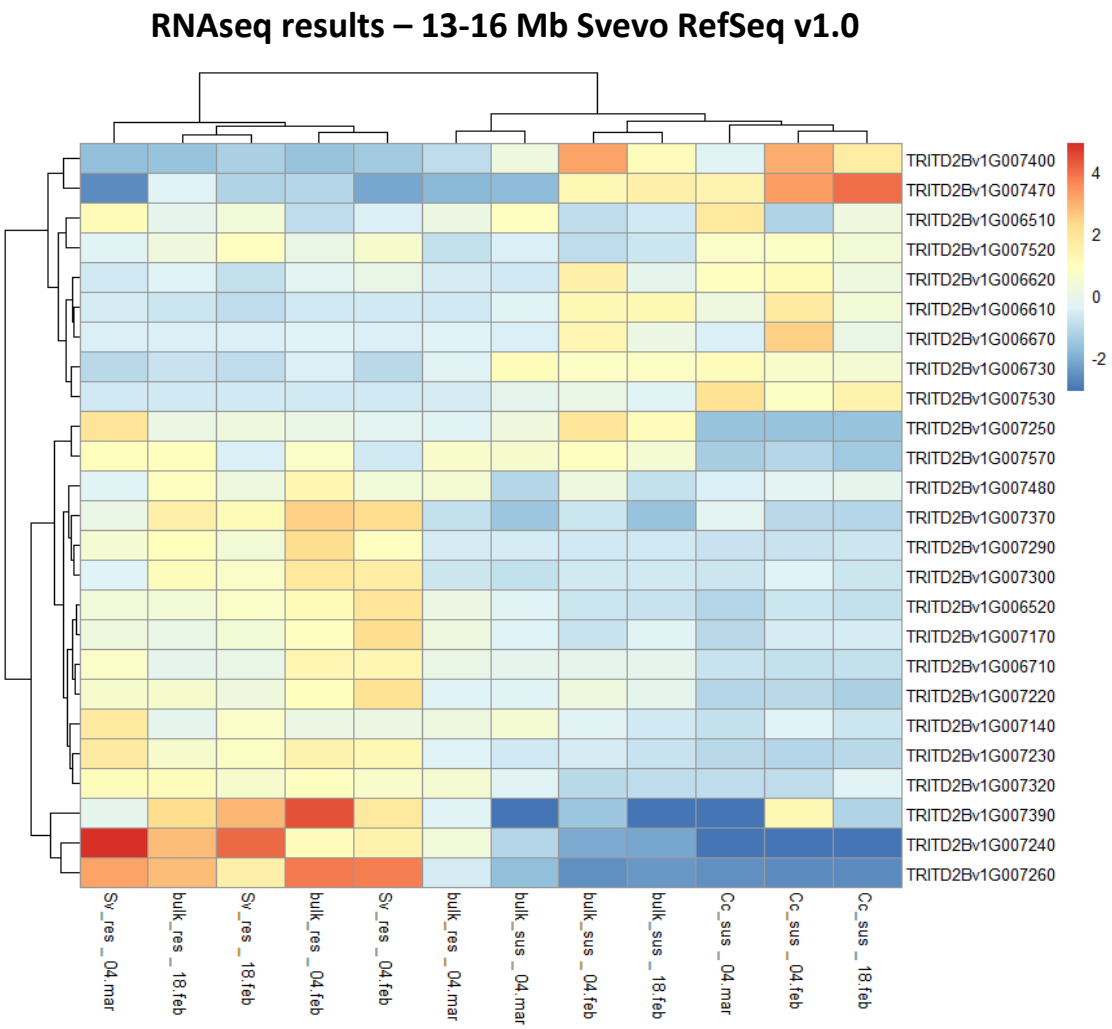


Figure 53: RNAseq results for the candidate genes within the sbm2 interval on Svevo RefSeq v1.0. Statistically significant differential expressed genes are shown in TPM values.

The RNAseq experiment clearly show that there is differential expression between candidate genes within the *sbm2* interval, especially regarding the NBS-LRR coding genes and protein kinases. Furthermore, Figure 53 shows that the expression level seems to be similar at the first two dates (4-02-2021 and 18-02-2021) whilst it is lower and the differential expression is less visible at 4-03-2021, especially between RB and SB. The hypothesis that seems valuable is that the resistance reaction is more effective during the replication of the SBCMV and *Polymixa graminis*. In fact, at 04-03-2021, viral symptoms were already visible in the field so, probably, the resistance reaction should occur earlier in order to hamper the spreading of the viral proteins into the plants and the symptoms to be visible.

The Svevo RefSeq v1.0 annotation was shifted to the Svevo Platinum pseudomolecule using the softwares *gmap* (Wu and Watanabe, 2005), and *blastn* to correct the ambiguous annotations for 900 genes. In addition, the final gff file with the shifted annotation was used to extract gene sequences and correct the alignment of the RNAseq reads against Svevo Platinum pseudomolecule using the software HISAT2.

In order to better understand if the resistance reaction occurs in durum wheat by the recognition of the vector *Polymixa graminis* or the SBCMV viral particles, the viral SBCMV RNA 1-2 and vector *Polymixa betae* public available genomes (same genus as *Polymixa graminis*) were used. The

unmapped reads from the alignment of RNAseq data against Svevo Platinum were mapped at first against the *Polymixa betae* reference genome (Decroës et al., 2019). A minimum part of the reads, between 500 and 600 reads, mapped against the vector genome without significant differences between resistant and susceptible samples (Figure 54).

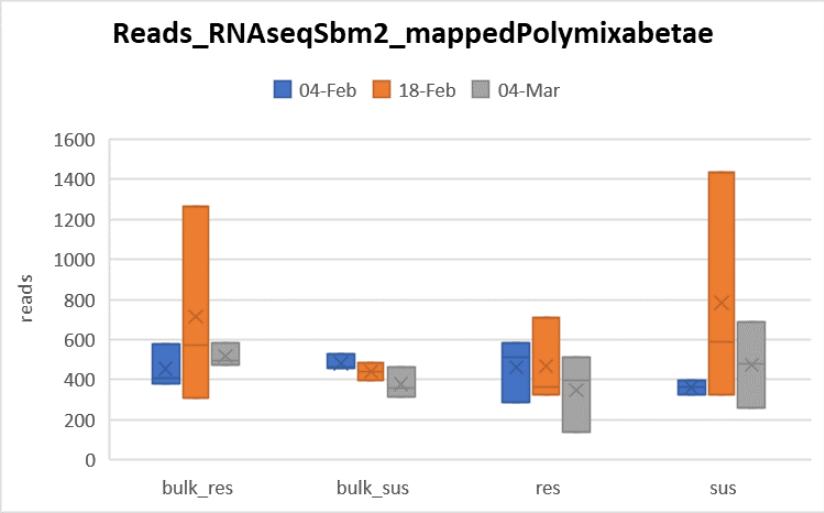


Figure 54: Unmapped reads against Polymixa betae genome for the following samples: resistant bulk (bulk_res), susceptible bulk (bulk_sus), resistant varieties (res) and susceptible varieties (sus).

On the other hand, about 35000 reads from the susceptible samples mapped against the RNA 1-2 of SBCMV, comparing to the resistant samples where about 100 reads mapped against the virus (Figure 55).

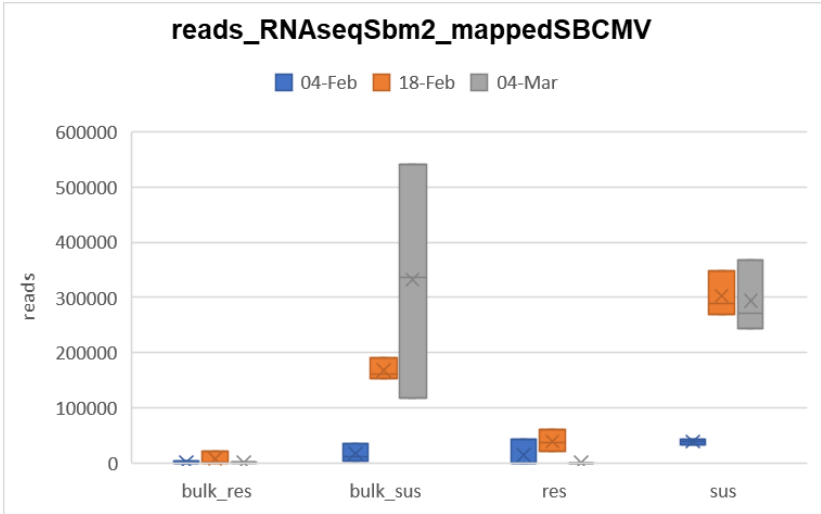


Figure 55: Unmapped reads against RNA1-2 of SBCMV from the following samples: resistant bulk (bulk_res), susceptible bulk (bulk_sus), resistant varieties (res) and susceptible varieties (sus).

The data show that there is a higher percentage of unmapped reads from susceptible samples in comparison to the resistant ones, when the alignment is performed directly on RNA1-2 of SBCMV. However, there is not significant difference when the mapping occurs on the *Polymixa betae*

genome that, even if it is not the specific *Polymixa graminis* genome, it is supposed that all the different genomes from this same genus could share a major part of the characteristics and not being very different one another. This may lead to the conclusion that the unmapped reads of RNAseq samples against Svevo Platinum pseudomolecule, could belong to SBCMV, hypothesizing that the resistant reaction occurs by recognition of the virus and not the fungal vector.
The RNAseq analysis was performed also against Svevo Platinum pseudomolecule, after lifting the annotation from Svevo RefSeq v1.0. Before proceeding with the final analysis, the RNAseq reads were aligned with HISAT2 on Svevo Platinum and transcripts in the *sbm2* regions were annotated *de novo* using Stringtie software, in order to check if there were other genes in the interval that could be involved in resistant reaction. The *de novo* annotation didn't show any error or difference from the reference gene annotation, with only genes *TRITD2Bv1G007290* and *TRITD2Bv1G007270* that showed different orientation and intron/exon overlap.
The final analysis was performed on Svevo Platinum liftover transcriptome annotation file using software kallisto and R package sleuth (Figure 56).

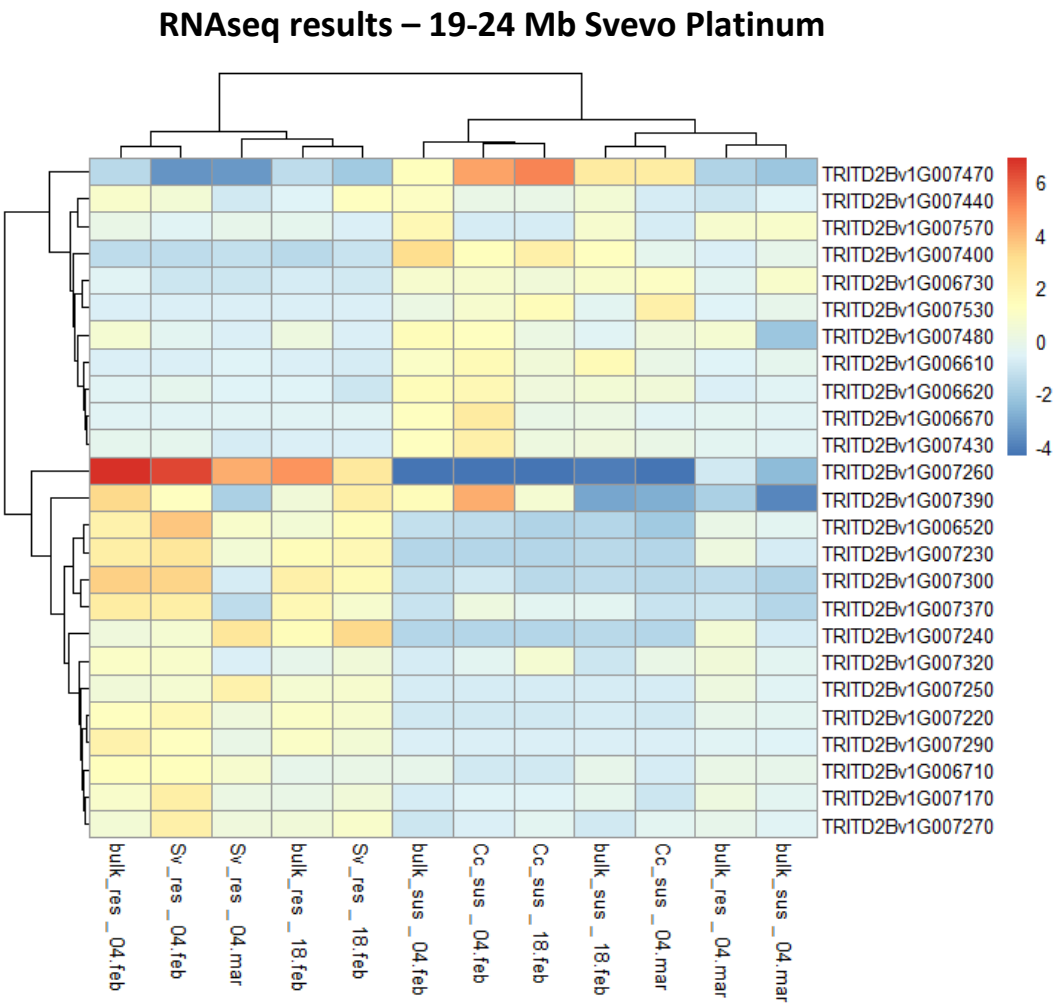


Figure 56: RNAseq results for the candidate genes within the sbm2 interval on Svevo Platinum, after the annotation liftover from Svevo RefSeq v1.0. Statistically significant differential expressed genes are shown in tpm values.

As reported in Figure 56 the resistant and susceptible clusters are still visible, but there are two genes strongly differentially expressed in resistant cultivars in comparison to susceptible ones, namely: *TRITD2Bv1G007260* and *TRITD2Bv1G007390*, two protein kinases. This confirms the

importance of these genes to be evaluated as candidates and also that, especially for the bulk samples, the differential expression on the 04-03-2021 is less evident. This strengthen the hypothesis that the resistance reaction in roots may occur during the viral replication phase, and not when the symptoms are already visible.

The gene network of the candidate genes *TRITD2Bv1G007260, TRITD2Bv1G007390* and *TRITD2Bv1G007240* was explored with Knetminer, using *Triticum aestivum* cv. Chinese Spring as orthologue. The gene network shows the molecular functions in which the genes are involved. It is worth pointing out that the gene *TraesCS2B02G033600* is the orthologue of *TRITD2Bv1G007260* in durum wheat and codifies for an NBS-LRR protein. In fact, the biological and molecular functions reported in the gene network correspond to defense response to virus, cell wall organization and other aspecific plant defense responses (Figure 57).

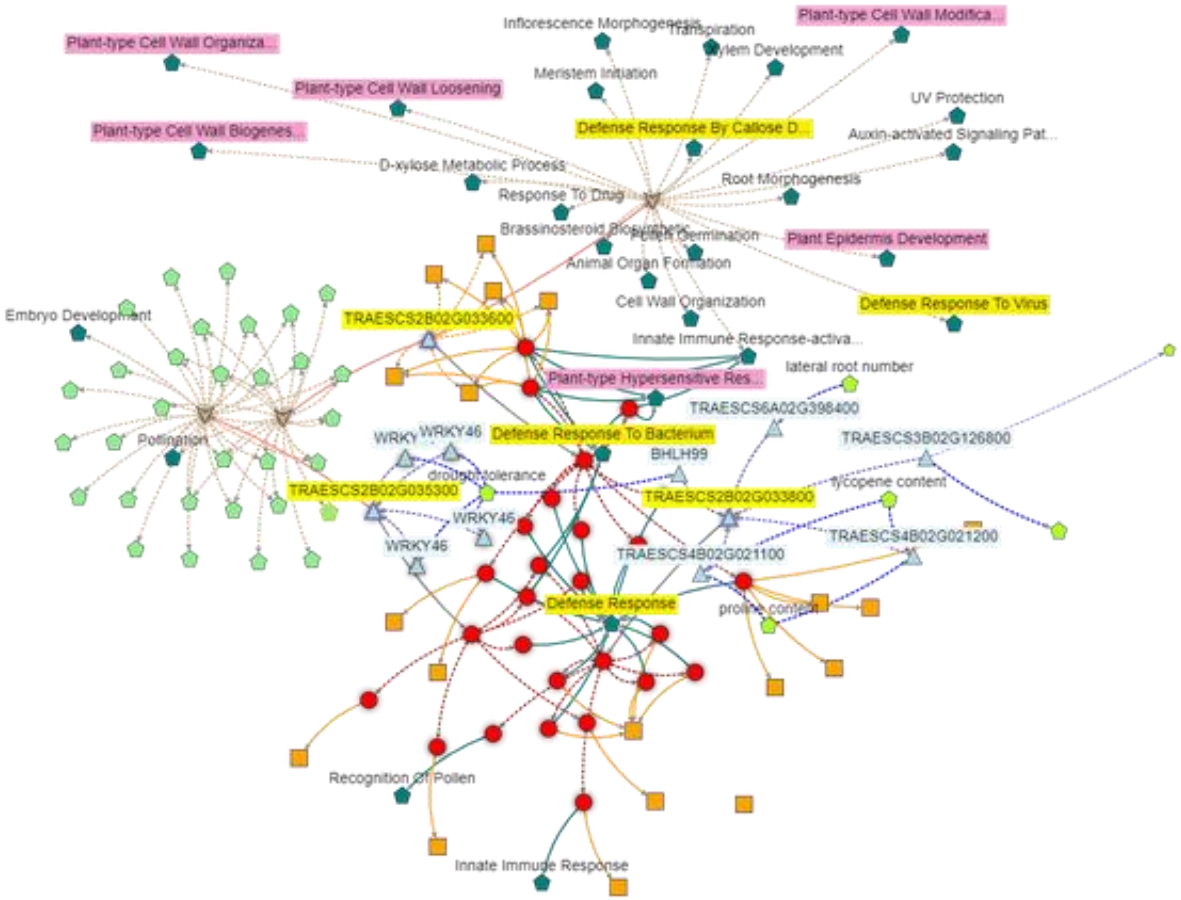


Figure 57: Knetminer gene network SBCMV resistance considering the candidate genes strongly differentially expressed between resistant and susceptible cultivars. Gene networks are involved in aspecific plant defense pathways, such as cel wall reorganization, brassinosteroid pathway and defense response to viruses. The light green pentagons and triangles represent the corresponding phenotype and trait of interest, the light blue triangles represent the genes involved in the network, the dark green pentagons report the biological process, the red circles report the associated proteins and the orange squares the related publications. Highlighted in yellow are reported the candidate genes with the main molecular functions.

5.3.6 *Sbm2* haplotypes based on durum panel

A haplotype analysis was performed on durum panel 1 (DP1) accessions, ca. 180 genotypes, on the *sbm2* QTL interval. The haplotype analysis identified five haplotype blocks where the fifth represented the main blocks that included the candidate genes within the confidence interval. This block included four different haplotypes, present among durum panel accessions for *sbm2* (Figure 58). Two out of the four different haplotypes, defined as haplotype one (the most frequent) and haplotype four (the least frequent) are associated with resistance based on adjusted visual score (VS) data (Figure 58 A) and adjusted ELISA data (Figure 58 B). The other two haplotypes, haplotype two and haplotype three, are associated with susceptibility based on adjusted VS data (Figure 58 A) and adjusted ELISA data (Figure 58 B) from DP1 (background data in Unibo).

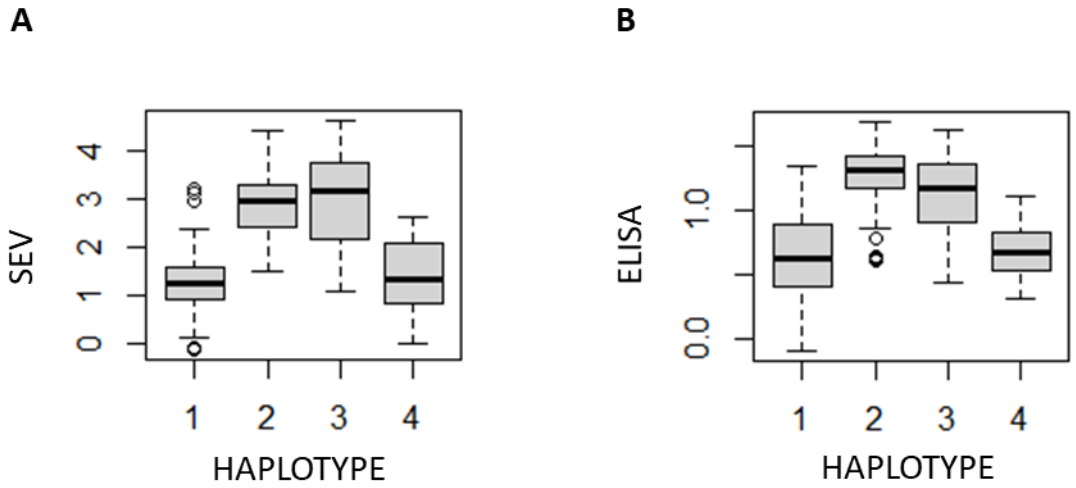


Figure 58: box plot showing the distribution of phenotypic values (y-axis) for Durum Panel accessions, based on the four haplotypes (x-axis) at block 5. Phenotypic values are BLUEs of symptom severity scores (SS; A) and ELISA (B) data collected in 2005 (VS), 2007 (SS and ELISA) and 2010 (VS and ELISA) in Cadriano (Bologna, Italy) and spatially corrected according to a moving average model based on population means considering the double of adjacent plots for each area. Haplotypes one and four are associated with resistance and haplotypes two and three are associated with susceptibility.

In order to broaden the field of investigation regarding *sbm2* haplotypes in durum wheat, a larger panel was investigated including the durum panel for Innovar collection. The total durum augmented panel included 549 accessions, including both DP1 and Innovar durum panel, were genotyped with markers KUBO 27, KUBO 29 and KUBO 1. Markers were chosen based on their position, with KUBO 27 and KUBO 1 being the flanking markers of the shortest interval determined by Sv x Cc fine mapping and KUBO 29 being inside the interval (Bruschi et al., *unpublished*). Moreover, these three markers showed to be efficient when tested on the DP1 (lowest level of conflicts).

Based on the genotyping results, haplotypes were classified in seven categories: resistant haplotype, susceptible haplotype, recombinant haplotype 1, recombinant haplotype 2, recombinant haplotype 3, double recombinant haplotype and admixed (this latter being defined as any haplotype different from the previous categories, including also missing data and heterozygous genotypes).

The proportion of accessions presenting each one of these haplotypes categories is reported in Figure 59. The resistant haplotype was the most present, representing the 53% of the accessions, followed by the susceptible haplotype, which represented the 31% of the accessions. The three

categories of recombinant haplotypes 1, 2 and 3 represented in total the 11%, while accessions with a double recombinant haplotype were less than 1%.
The distribution of resistant, susceptible and recombinant (1, 2 and 3) haplotypes based on varieties' origin is reported in Figure 60. For all origin locations, varieties with a resistant haplotype represented more than 45% of the total, with Portugal-Spain having the minor proportion of resistant accessions (46%). The highest proportion of resistant varieties was reported in Australia (100%) even though only three Australian cultivars were present in the panel. In the North American continent, Canada-USA beat CIMMYT-Mexico with 81% versus 49% of resistant varieties. In Europe the resistant haplotype was most represented in France (51%), even if the proportion was very similar to that of the other European countries, ranging from 46% to 51%. Resistant varieties from ICARDA were 72%. The susceptible haplotype was totally absent in the Australian varieties and poorly represented in Canada-USA locations (14%) and ICARDA (25%), while ranged from 35% to 44% in all other locations. The proportion of recombinant varieties was the least represented ranging from 0% (Australia) to 16% (CIMMYT).
Analysing haplotypes trends based on the year of release of varieties from 1955 to 2020 it is worth to note that the release of varieties in the North American continent has undergone an outburst in the interval 1995 - 2000: while in Canada - USA the great majority of released varieties have the resistant haplotype, varieties from CIMMYT – Mexico are more balanced in terms of proportion of resistant/susceptible haplotypes. In Italy and France, the release of varieties has increased starting from 2005 and the proportion of resistant haplotypes had been higher starting from that year in respect to the susceptible haplotype. In the Central – Eastern Europe the initial release of most resistant varieties left the place to the release of susceptible varieties starting from 2000. For what concerns Portugal-Spain the release of varieties is more constant throughout the years, with a major release of resistant varieties in the interval 1995 – 2000. Release of varieties from ICARDA was concentrated in the first half of the considered period, from 1985 to 2000 and the majority of varieties presented the resistant haplotype (Figure 61).

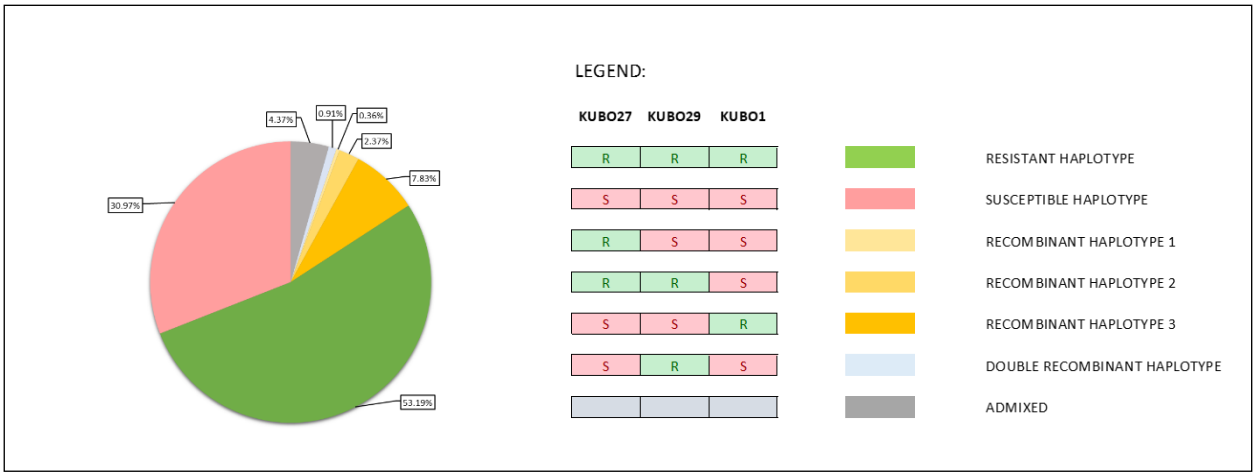


Figure 59: Proportion of different haplotypes at sbm2 composing the panel of 549 durum wheat accessions. The haplotype refers to three markers (KUBO 27, KUBO 29 and KUBO 1) and includes the support interval of the QTL.

Figure 60: Proportion of resistant, susceptible and recombinant haplotype in durum wheat varieties based on their origin. The number in brackets refers to the total number of accessions. The haplotype refers to three markers (KUBO 27, KUBO 29 and KUBO 1) and includes the support interval of sbm2.

Figure 61: Histograms showing Qsbm.ubo-2BS haplotypes trends across intervals of 5 years based on the years of release of the accessions and on their origin. The haplotype refers to three markers (KUBO 27, KUBO 29 and KUBO 1) and includes the support interval of sbm2. Two accessions of Central – Eastern Europe, respectively resistant and released in 1926 and recombinant and released in 1950, and three accessions of Italian origin, respectively recombinant and released in 1915 and susceptible and released in 1930 and 1940, are not reported in the histograms.

5.4 Discussion

The s*bm2* QTL responsible for resistance to SBCMV was first detected in durum wheat on the short arm of chromosome 2B using mapping RILs population Meridiano x Claudio and Simeto x Levante (Maccaferri et al., 2011b). In addition, this QTL was confirmed by Bruschi et al. (2022, unpublished) based on QTL mapping on RILs population Svevo x Ciccio and Meridiano x Claudio, together with a GWAS analysis on durum panel 1 (Maccaferri et al., 2015). The *sbm2* QTL was also detected in bread wheat on the same chromosome, but an additional QTL was detected on long arm of chromosome 5D (*sbm1*), responsible for wide spectrum resistance to different races of soil borne viruses (Bass et al., 2006; Kanyuka et al., 2004).

The herein work reports the fine mapping procedure for the *sbm2* QTL exploiting two genotyping platforms, Infinium SNP 90K Chip array and Axiom Affimetrix 420K SNP Chip. Basically, about 20 KASP markers were converted from the genotyping platforms included in the *sbm2* QTL interval of 2 cM. Based on the analysis on the RILs mapping population Meridiano x Claudio and backcross Meridiano x MC F_4 the s*bm2* interval was fine mapped to a region of 1.1 Mb on chromosome 2B, between 14.7 Mb and 15.8 Mb flanked by the markers KUBO 63/76 and KUBO29/81. The interval was confirmed by multi-years field trial where symptoms severity was evaluated by visual score (0 as resistance and 5 as susceptible) and by ELISA molecular test.

This fine-mapping procedure using KASP markers has been extensively used form different studies to narrow down QTLs interval, such as studies regarding main agronomic traits and disease resistance traits (Rahman et al., 2020; Dong et al., 2022; Duan et al., 2022). The KASP technology is very efficient in detecting the SNP marker segregation for the QTL of interests, that can be developed from different SNP Chip arrays. For example, the procedure followed by Rahman et al., (2020) is similar to the one developed herein for *sbm2* fine mapping as, after QTL detection, KASP markers were used to fine map the QTL confidence interval to a smaller genomic region to detect the most probable candidate genes.

The fine-mapped interval contains about 40 genes between high confidence and low confidence, whose candidates can be grouped in few categories: cytochrome p450, defensins, protein kinases and NBS-LRR coding genes. The last two categories can be ascribed as the principal candidates for the resistance, as they are involved in signal transduction, wide spectrum defense and effector triggered immunity (ETI) (Dodds and Rathjen, 2010). The ETI immunity system is based on the direct recognition of NBS-LRR proteins with the avirulent protein from the pathogen. The recognition can occur by different types of interaction mechanism, the clusterization of the NBS-LRR protein coding genes in the *sbm2* interval can hypothesize also interaction between different proteins, such as prey-bait mechanism of avirulence protein recognition (Marchal et al., 2020).

The genes transcription level was explored by an RNAseq experiment on susceptible and resistant durum wheat varieties. The varieties were sown in a homogeneous SBCMV infected field trial and root samples for RNA were collected at different dates during the growing season, corresponding to different phases of viral, vector development and differentiation. Based on the RNAseq results, differentially expressed genes were identified, whose expression can be grouped in clusters. In fact, a group of genes is more expressed in resistant samples and another cluster is more expressed in susceptible samples. The genes are clustered based on Ward algorithm, showing small cluster of three genes strongly overexpressed in resistant cultivars, two of which are protein kinases and one is a NBS-LRR protein coding gene. The main consideration is that none of the genes differentially expressed across the clusters can be excluded as candidate for the resistance reaction, but priority can be given to protein kinases and NBS-LRR proteins. The latter usually cluster together in close genomic regions, as shown using NLR-annotator software (Steuernagel et al., 2020).

Furthermore, the new version of Svevo assembly was available in Unibo, namely Svevo Platinum (Maccaferri et al., *unpublished*). The *sbm2* interval was compared, shifting the annotation of Svevo RefSeq v1.0 to Svevo Platinum and reannotating de novo transcript included in the *sbm2* region. The comparison between the two intervals shows an inversion between Svevo RefSeq v1.0 and Svevo Platinum which is included in the interval reported by Bruschi et al. (*unpublished*) but it brings new genes to be considered focusing on the fine mapped interval (17.8 Mb – 15.8 Mb), that are involved in cytochrome p450. However, no significant differences in the annotation were identified between Svevo RefSeq v1.0 and Svevo Platinum and the cytochrome p450 genes, included in the fine mapped interval of the correct assembly Svevo Platinum, were not differentially expresses re-analysing the RNAseq data on Svevo Platinum transcriptome. The fact that the comparisons between the two assemblies revealed some differences in the orientation of chromosome regions was not totally unexpected, as it may depends on differences of sequency technology (Illumina HiFi for Svevo RefSeq v1.0 and PacBio for Svevo Platinum) and contigs length that could lead to errors during the pseudomolecule assembly. The fact that PacBio technology was used to sequence the new version of Svevo Platinum pseudomolecule gives more reliability to this assembly in comparison to the Svevo RefSeq v1.0. Similar tools used to check differences between Svevo RefSeq v1.0 to Svevo Platinum were used also to compare the *Triticum aestivum* cv. Renan assembly to the published ten bread wheat genomes (Walkowiak et al., 2020; Aury et al., 2022). As shown also by the Svevo assemblies' comparisons, there is very little difference between the assemblies with few inversions between Renan and all the other assemblies, using similar parameters in the sequence comparison analysis to the ones reported in the current thesis chapter.

The RNAseq analysis using Svevo Platinum as reference showed similar results to the previous outcomes on Svevo RefSeq v1.0. In fact, same differentially gene expression clusters were detected with the most differentially expressed genes involved in NBS-LRRs (*TRITD2Bv1G007230* and *TRITD2Bv1G007240)* and protein kinases (*TRITD2Bv1G007260* and *TRITD2Bv1G007390).*

The final question was to determine who is recognized for the resistance reaction: the fungal vector *Polymixa graminis* or SBCMV. Based on literature research, the real trigger for the resistance reaction is still uncertain, different methods were published to determine the presence of viral proteins inside the host plant (Kanyuka et al., 2004; Ratti et al., 2004) but the interactors are not known. RNAseq unmapped reads were used to try to answer to this question, finding that an high percentage of unmapped reads is on Svevo Platinum were mapped in SBCMV RNA 1-2 in susceptible samples in comparison to the resistant ones. No differences were detected between resistant and susceptible samples mapping the reads on the available genome of a close fungus relative, *Polymixa beate*. This shows that there is no presence of vector genes inside the host plants, that could lead to a recognition to the resistance genes. On the other hand, the susceptible samples host an high percentage of viral genes, which seems not to be recognized by the resistance genes allowing the virus to replicate inside the host plant.

To conclude, the herein reported work show the fine mapping procedure of the *sbm2* QTL in durum wheat with the characterization of the candidate genes involved inside the interval. Furthermore, the haplotypic analysis on the durum panel 1 augmented with Innovar durum panel, show that sources of resistant haplotypes gradually decrease in recently registered varieties. The *sbm2* QTLs is less spread in the analyzed germplasm, composed of modern registered varieties commercially distributed in European market (Innovar germplasm), probably because, as shown by ancestry analysis, the major representation of breeding programs comes from ICARDA and CYMMIT, where the presence of SBCMV is not diffused. On the other hand, the *sbm1* QTL (Bass et al., 2006), mapped on chromosome 5DL of bread wheat, is more stable at high frequencies in bread wheat modern germplasm, especially in UK germplasm. The link between durum and bread wheat regarding SBCMV resistance is not fully understood and explored. *Sbm1* and *sbm2* are two separate QTLs for

the same resistance reaction to SBCMV, however it is known that some bread wheat varieties have both QTLs (es. Cadenza and Cordiale), whose mapping population with susceptible parental lines are under evaluation by Unibo and other partners (Kanyuka et al., 2004).
The fact that *sbm2* is not spread in modern varieties reflect the lack of genetic diversity and fixation of different alleles of recent breeding programs. In the past, different durum varieties were identified as sources of resistance to SBCMV, such as Meridiano, Svevo, Levante, Neodur and few others mainly registered during 90's or earlier (Maccaferri et al., 2012, 2011b; Ratti et al., 2006; Russo et al., 2012; Vallega, 1985). This strengthens the importance of detecting the candidate genes and piramidize the QTLs both in bread and durum wheat varieties under registration procedure, trying to counteract the SBCMV spread that depletes yield production about 70% in Europe.

6. Characterization of the *GNI-2A* QTLs in a biparental durum population

6.1 Introduction

6.1.1 Yield related traits in wheat

Yield is a very complex trait under the regulation of multiple QTLs (Sakuma and Schnurbusch, 2020) in all grasses, wheat in particular. This agronomic character is strongly linked to the environment and is dependent by different phenological features (Reynolds et al., 2001). Two equations explain the components that play a role in the yield determination (González et al., 2003; Hay, 1999):

GY=BY * HI developed in GY= Q * I * RUE * HI

where GY, BY, HI, Q, I, RUE, stand for: grain yield, biomass yield, harvest index, amount of incident radiation during growing season, the fraction of incident radiation intercepted by the crop canopy and the efficiency of the crop to convert this radiant energy into dry matter, respectively.
The final yield of a plot is calculated using the following formula:

GY=NG m^{-2} * IGWt

Where NGm^{-2} stands for number of grains per square meter and IGWt stands for individual grain weight.
Other components of yield could be calculated after plot harvest, such as: number of spikes/m^2, number of grains/spike and thousand grain weight (TGW). The latter has a lower range of value than NGm^{-2} and a minor correlation with total yield (Slafer and Rawson, 1994; Slafer and Andrade, 1993). As a consequence of climate change and increase in world population size, there is the need for agriculture to focus on grass productivity and sustainability (Sakuma and Schnurbusch, 2020). Because of that, the interest in increasing grass yield and grain number is of major importance for breeders and genetic programs (Lynch et al., 2017; Voss-Fels et al., 2019; Würschum et al., 2018).
Grains of wheat (both durum and common) are located in single florets organized in spikelets. The spikelets are arranged on a stem called rachis alternating on opposite sites symmetrically (Bonnett, 1966; Newton et al., 2011; Shiferaw et al., 2013). The total number of spikelets composes the spike itself, with a terminal spikelet at the top of it (Bonnett, 1966). Each spikelet is organized in an undetermined number of florets placed on a secondary axis called rachilla (Figure 62). The number of florets per spikelet is directly connected with the grain number trait the spike fertility (Gauley and Boden, 2019; Sakuma et al., 2019; Sakuma and Schnurbusch, 2020).

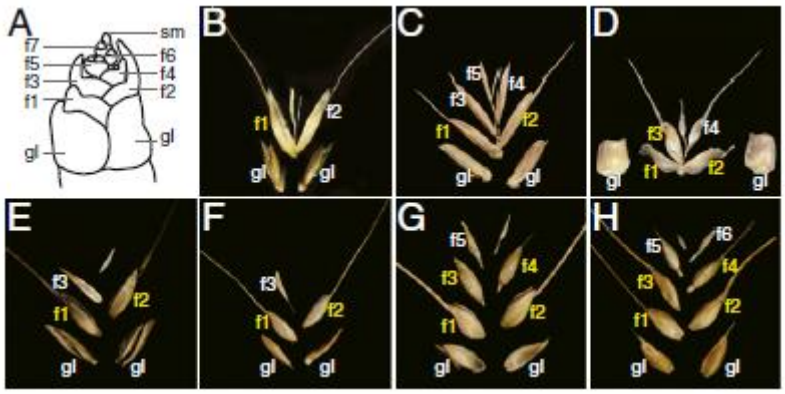


Figure 62: spike floret fertility architecture in bread and durum wheat, where gl represents the glumes, f the number of florets and sm the spikelet meristem. A) reports a spike at immature stage, B) Triticum Urartu, C) Aegilops speltoides, D) Aegilops tauschii, E) wild emmer, F) domesticated emmer, G) durum wheat H hexaploidy bread wheat (Sakuma et al., 2019).

As already mentioned, yield is a complex trait under control of different agronomical features, but its complete regulation is still under investigation. For example, domestication played a role in inflorescence development and morphology (Doebley et al., 2006). One of the main events was the gain of non-threshing which prevented the grain to fall to the soil at complete maturity, controlled by few genes such *as Non-brittle rachis 1 (btr1)* (Avni et al., 2017; Pourkheirandish et al., 2018, 2015) and *Shattering1 (Sh1)* (Lin et al., 2012). After fixation of the non-brittle rachis allele, breeding selection for increased yield favoured unconsciously mutants with increased fertility in spikelets, converting sterile florets to fertile florets (Komatsuda et al., 2007; Sakuma et al., 2019).
In Nature, most of phenotypically active genes are transcription factors that respond to environmental and endogenous stimulus to fine tune/regulate the functioning of structural genes. Among transcription factors, we have activators and repressors.
The study of grasses inflorescence is focused on understanding floret sterility as a consequence of strong repressors naturally selected through millions of years and the conversion from sterile florets to fertile in order to increase spike productivity. As regards to barley inflorescence, *Vrs1* (HD-Zip I transcription factor) gene was first identified to have a contribution in grain number increase in barley six-rowed spike, by converting sterile to fertile lateral florets(Komatsuda et al., 2007; Sakuma et al., 2013). In addition to *vrs1*, the *Vrs5* gene was identified as the orthologue of maize domestication gene teosinte branched 1 (*tb1*) (Lundqvist and Lundqvist, 1988; Ramsay et al., 2011). The *Vrs5* loss of function alleles confers high grain number in lateral spikelet and increased number of tillers in six row barley, the gene encodes for a transcription factor. Other genes involved in grain number for lateral spikelet development in six-rowed barley are *Vrs4*, which regulates lateral spikelet fertility in spikelet primordia (Koppolu et al., 2013), and *Vrs3* which promotes the activation of *Vrs1* acting as a positive regulator (Bull et al., 2017). Taken together, these findings indicate that *Vrs3* and *Vrs4* regulate row-type architecture of the barley inflorescence by converging to positively regulate transcription of *Vrs1*. The last identified gene was *Vrs2* which encodes for *SHORT INTERNODES (SHI)* transcriptional regulator during barley inflorescent development, promoting two-rowed spikelet architecture by regulating hormone level (Youssef et al., 2017). Figure 63 reports the actions of the *Vrs* genes in the floret's determination and lateral spikelet development in barley.

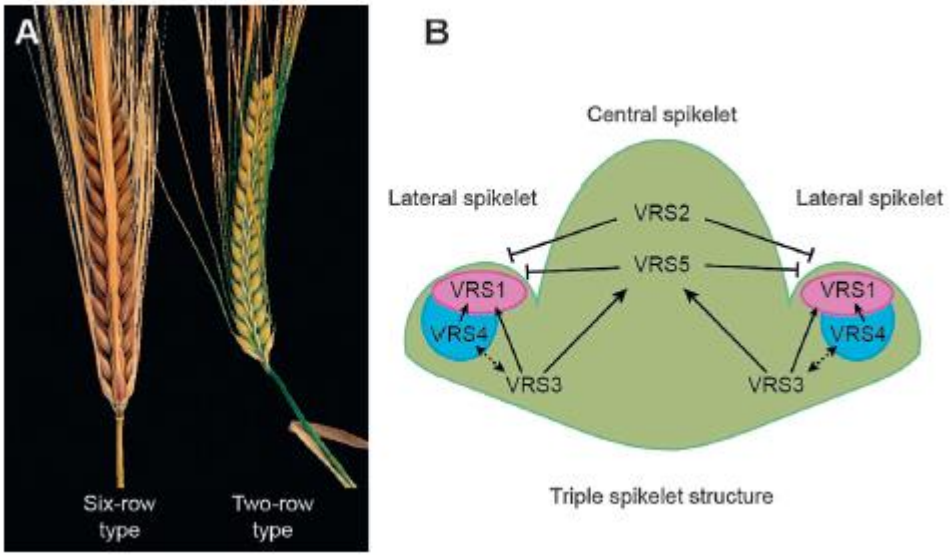


Figure 63: A) spike morphology of two and six row barley spikes. B) Vrs genes schematic activity where pink and blue spots reports the site of expressions and arrows report the positive or negative effects of Vrs2,3,5 on Vrs1,4 (Boden et al., 2019).

The studies carried out in barley provide several access points for wheat, also in consideration of the extended synteny and conservation of the gene functions between the two species. As regards to regulation of floret abortion, *Grain number increase 1* (*GNI1*), in wheat, encodes for a transcription (*HD-Zip 1*) factor responsible for abortion, orthologue of *Vrs1* in barley (Sakuma et al., 2019) (Figure 64). *GNI1* expression is detectable in apical florets primordia and in part on the rachilla, where it inhibits floret development. It has been shown that its loss of function, due to single amino acid substitution (N105Y) in the conserved domain, enhance the number of fertile florets per spike. The duplication of *GNI1* in *Triticeae* genus with mutation increasing floret fertility have been under selection during domestication (Sakuma et al., 2019).
The *GNI1* paralogue, *hox2*, is a gene conserved among grasses which codifies for the same class of transcription factors as *GNI1*. Its overexpression in bread wheat produces spikes with a lower number of spikelets and decrease in grain numbers (Wang et al., 2017).

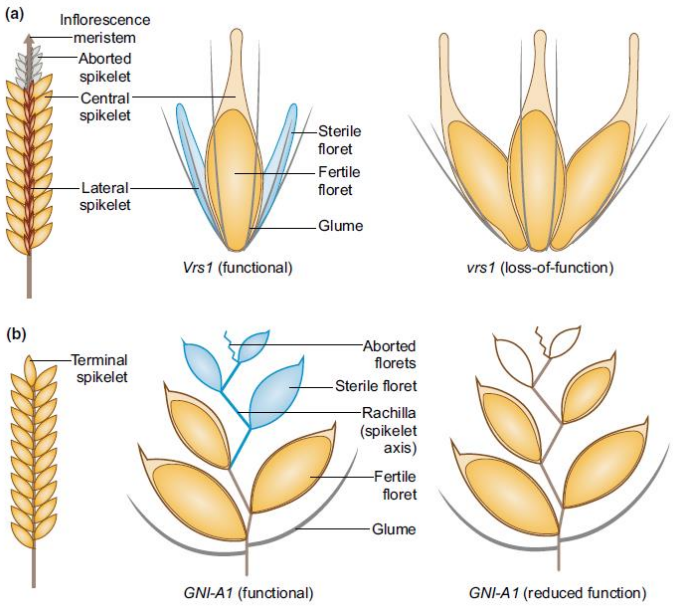


Figure 64: structure of barley and wheat inflorescence reporting the functions of vrs1 and GNI1 in wt or mutated form (Sakuma and Schnurbusch, 2020).

Beside the controls of the number of florets per spikelets, another important trait is the flowering time that are influenced by temperature and photoperiod. The flowering time is controlled by the *Flowering locus T (FT)* (Kardailsky et al., 1999; Kobayashi et al., 1999; Turck et al., 2008). This gene promotes flowering based on the photoperiod activating expression of meristem genes in the shoot apical meristem (Boden et al., 2015; Corbesier et al., 2007; Tamaki et al., 2007). This gene is regulated by the *Photoperiod-1* (*Ppd-1)* which regulates photoperiod responsive flowering pathways in wheat and barley. The gain of function of *Ppd-1* reduce the number of spikelets per spike caused by a reduce duration of early developmental stages (Ochagavía et al., 2018). The *FT-1* copy on the B genome has a role in the number of spikelets, acting downstream of *Ppd-1*, *FT-B1* mutants show an increase in spikelet number (Dixon et al., 2018; Finnegan et al., 2018).
These results show how the inflorescent architecture is regulated, both in barley and wheat, by different transcriptional and post-transcriptional regulation of flowering signals that regulate the inflorescent architecture. The spike architecture is regulated by different genes that act at different steps of inflorescent development and interact each other, coordinated also by the photoperiod and temperature of environment.

6.2 Materials and methods

6.2.1 Background material in UNIBO

This project is in collaboration with IPK (Gatersleben, Germany), where background activities and material were developed to understand the genetic regulation of grain number increase in central spikelets of durum wheat and of spike fertility in general. The main results of this collaboration were reported in the paper by Milner et al (2016). Briefly, a four-way population composed by Neodur, Claudio, Colosseo and Rascon2/Tarro (NCCR) was developed and phenotyped for different yield traits (Milner et al., 2016), used for association mapping analysis. Neodur is a French photoperiod-sensitive late cultivar showing high number of spikelets per ear; Claudio shows wide adaptability to Southern Europe and resistance to drought and powdery mildew; Colosseo presents high-yielding ears (with balanced yield components); Rascon/2*Tarro (related to Altar-84) is a photoperiod-insensitive cultivar with high yield potential due to a high number of grains per spikelet.
Briefly, based on association mapping on NCCR population, *QGns.ubo2A* QTL (chr2A, QTL s.i. 84.6-87.6 cM) QTL, here referred as *GNI-2A*, was detected as strongly involved in phenotypic variation for grain number per spikelet increase (GNI). Comparing different haplotypes of the NCCR population based on Illumina iSelect Infinium SNP 90K chip, the Rascon/2*Tarro haplotype effect was responsible of an increase of +0.55 grains per spikelet. On the other hand, all the three other parents show a negative GNI effect at the locus. Furthermore, it was identified that the Rascon/2*Tarro haplotype was consistent with the Altar-84 haplotype, historic parental varieties from CYMMIT germplasm, and all the related varieties. Starting from the identified QTL peak on the 2A chromosome, a candidate gene was detected with a collaboration between Unibo and IPK, from a blast analysis against emmer and durum wheat assemblies the candidate gene seems to be *hox2,* which codifies for an Homeobox-Leucin Zipper transcription factor (HD-Zip). The gene presents a deletion of 4kbp in Altar84 haplotype related variants (such as Rascon/2*Tarro, Saragolla and Iride), highly correlated with increase in numbers of fertile florets per central spikelets.

6.2.2 Plant material

The haplotype analysis with *GNI-2A* KASP markers was performed on the durum panel 1 (DP1) on 167 genotypes (Maccaferri et al., 2015).
The fine mapping of *GNI-2A* QTL occurred on a biparental population acquired in F_3 from Florimond Desprez (France) and obtained from a biparental cross between Relief and Iride. Relief is a French variety characterized by a long spike and awns and with late flowering time, Iride is characterized by short spike with increased florets per spikelet, related to Altar-84 haplotype on *GNI-2A* QTL, which is the CYMMIT variety first identified to carry the grain number increased (GNI) phenotype.
The population (1500 genotypes) was multiplied in Cadriano (Unibo) field station (44°330N lat., 11°240E long., Bologna, Italy), during 2018 in 0.5x1m plots, and sown for fertility trait evaluation in two consecutive years, 2019 and 2020, reaching F_5 and F_6 generation. These field trials were conducted in complete randomized block design with two replicates using common repeated controls, namely: Relief, Iride, Altar-84 and Svevo.
Genotypes were harvested at complete maturity collecting 6 separate plants per genotype, main culms were collected for phenotypic evaluation and bulk seeds were threshed for all the secondary culms.

6.2.3 Phenotypic analysis

Spike fertility trait was assessed for each genotype and replicate of Relief x Iride F_5/F_6 population on 6 main culms harvested from selected plants and from harvested spikes from NCCR multi-parental cross (average of six plants).
The following parameters were measured: spike length, number of fertile spikelets, number of unfertile spikelets, number of total florets per central spikelet and number of grains per central spikelet (Figure 65).
The central spikelet was considered as the 9^{th} or 10^{th} starting from the bottom of the spike, the bract (sterile floret) at the top of the spikelet was not included in floret counting.

ALL	average spike lenght
	spike_number
SPIKE_n	Sterile_spikelet_number
SPIKE_n	Fertile_spikelet_number
SPIKE_n	floret number per central spikelets
WHITE/BRONZE/BLACK	GLUME_COLOR
WHITE/BRONZE/BLACK	AWN_COLOR

Figure 65: Phenotypic tables to monitor spike fertility traits in durum wheat RILs population and panels. The total number of spikes analyzed per genotype was at least 6.

6.2.4 Statistical analysis on phenotypic data

Phenotypic data were statistically analysed using software R with packages *car*, *cardata*, *lme4* and *heritability*. Fertile florets per central spikelet were considered and outliers were removed using the interquartile rules: the interquartile range (IQR) of the data was multiplied by 1.5, outliers were defined as values 1.5 *IQR above the third quartiles and 1.5 *IQR below the third quartile. Based on shapiro-test analysis, not normal distributions were normalized using power transform algorithm, in order to be analyzed by ANOVA via linear mixed model analysis with the *lme4* package. Basically, the *lmer* (mixed model using *lme4* package) function was used to analysed phenotypic data, including environmental variables as covariates, such as rows, columns, replicates, number of plants considered and the interaction between number of plants within the genotype. From the mixed model, BLUEs were calculated using genotype as fixed and other parameters as random using the following formula:

Florets/central_spikelet ~ Genotype + Replicate + rows + columns + number_of_plants + Genotype:number_of_plants

Heritability was calculated using package *heritability* using RStudio software.

6.2.5 DNA extraction and sample preparation

A sub-selection of durum panel I (DP1, 167 genotypes) (Maccaferri et al., 2015) and Relief x Iride population (1500 genotypes) were sown to extract DNA and perform wet lab analysis. Briefly, 5 seeds from different selected spikes were sown in a multi-pot trial and grown for 10 days in greenhouse with 14h light, 25°C during day period and 16°C during night period. Once plants reached 10cm length, leaves were harvested and freeze dried using a lyophilizer for 72h. Subsequently, leaves were grinded in extraction plates and DNA was extracted using CTAB protocol (Doyle and Doyle 1987) with minor adjusted steps.
The DNA was quantified using both agarose gel at 1% concentration and biophotmeter, checking for high quality value of 260/280 and 260/230 ratios. DNA was diluted at a concentration of 25ng/µl for genotyping.

6.2.6 KASP genotyping

Illumina Infinium SNP 90K Chip genotyping array dataset (Maccaferri et al., 2015) was used to analyse the haplotype *GNI-2A* QTL on durum panel accessions. 90K SNP markers with contrasting alleles within *GNI-2A* chromosomic region between Svevo and Altar-84 were converted to KASP markers. KASP primers A and B were designed on the specific SNP being variety specific, whilst the common C primers was designed to be specific for the A genome on *Triticum turgidum* cv. Svevo. The KASP primers were developed following methodology reported in paragraph 5.2.4.
KASP primers specificity was tested on Relief, Iride, Svevo and Altar-84 durum genotypes. Once confirmed, genotyping was performed on durum panel 1 (Maccaferri et al., 2015) and Relief x Iride RILs F_6 using KASP genotyping kit (LGC genomics). A and B primers were attached with different tails complementary to different fluorescent probes (FAM and HEX) included in the master mix, to discriminate different allelic variants on specific SNPs: GAAGGTGACCAAGTTCATGCT (5'-FAM) for primer_A and GAAGGTCGGAGTCAACGGATT (5'-HEX) for primer_B (
Table 39). Primer C was common to both primer A and B and was designed on homoeologues SNP for genome A.

Table 39: KASP primers developed from Illumina Infinum 90K SNP Chip. The different target SNP in primer A and B are reported in capital letter.

SNP ID	KASP	Primer A	Primer B	Primer C
IWB24557	**KUBO44**	tgagttggattggatggttgA	tgagttggattggatggttgC	gctacctcaaggaaggttcttc
IWB45502	**KUBO45**	atgatgacacaacaccaggT	atgatgacacaacaccaggG	gaggcctgatcgttgctgt
IWB45503	**KUBO46**	cctctaaccgtcacaatgctA	cctctaaccgtcacaatgctC	tctcggcttgttcattgagtc
IWB63013	**KUBO47**	tctgtgcaagcgtgataagA	tctgtgcaagcgtgataagC	atctgagctgctcccacaa
IWB69369	**KUBO49**	gcagtgctgattgagacttacttaA	gcagtgctgattgagacttacttaC	cgaacttaggtctaggcctgg
IWA581	**KUBO50**	gggttcatactacaatggtggttT	gggttcatactacaatggtggttC	tgaactcaatatcgatcctgaatca

Additional KASP markers were converted from Affimetrix 420K in order to fine map the *GNI-2A* confidence interval (Table 40).
SNP sequences inside the confidence interval were aligned on 55 durum varieties from durum panel 1 (Maccaferri et al., 2015) to select the polymorphic probes on durum wheat, markers with contrasting alleles on selected cultivars were converted to KASP and tested to be specific as performed for the previous KASP markers.
KASP converted from Affimetrix 420K were tested on ten Relief x Iride selected recombinants lines.

Table 40: KASP primer A, B and C sequences converted from Affimetrix 420K SNP Chip. Position of each probe and KASP numbers are reported. A and B primers (primers forward 1 and 2) and common primer sequences are reported

SNP ID	KASP	position	Primer Forward 1	Primer Forward 2	Common Reverse
AX-89710414	**KUBO 172**	151074385	GAAGGTGACCAAGTTCATGCTAGTGCAAATAATGAAAGCAGTAAAAA	GAAGGTCGGAGTCAACGGATTAGTGCAAATAATGAAAGCAGTAAAAG	CGCCATTTCAATTTTCTACTTCTG
AX-89584287	**KUBO 173**	153155184	GAAGGTGACCAAGTTCATGCTATAAACTAAAGCAGCTGAAAGAAAAT	GAAGGTCGGAGTCAACGGATTATAAACTAAAGCAGCTGAAAGAAAAC	TATTATTGTCGCACTGGCACACC
AX-89584287	**KUBO 174**	153155184	GAAGGTGACCAAGTTCATGCTAACCTCTCTATTGTAGACAACAACCA	GAAGGTCGGAGTCAACGGATTACCTCTCTATTGTAGACAACAACCG	GTGAAAGAGTTGATGATTGCTGCA
AX-89623025	**KUBO 175**	153760282	GAAGGTGACCAAGTTCATGCTGAGTTAATAATTCAACGCATTTGATA	GAAGGTCGGAGTCAACGGATTGAGTTAATAATTCAACGCATTTGATG	ATTTCGCACTCATAAGCAGAATC
AX-89612309	**KUBO 176**	153574054	GAAGGTGACCAAGTTCATGCTAAACATCACTAGCCTTTATTCTCTCT	GAAGGTCGGAGTCAACGGATTAAACATCACTAGCCTTTATTCTCTCC	CAGGGACACTTTATAAGTACCCC
AX-89505095	**KUBO 177**	150351083	GAAGGTGACCAAGTTCATGCTAGGATTTGACACGAGCCTCGTT	GAAGGTCGGAGTCAACGGATTGGATTTGACACGAGCCTCGTC	CCAGTGAAACACCCCGAGTACTA
AX-89459527	**KUBO 178**	151565662	GAAGGTGACCAAGTTCATGCTATGCACGCGAGCCAGGAATCA	GAAGGTCGGAGTCAACGGATTCATGCACGCGAGCCAGGAATCG	GATCCCACCTCTGTTGGTGGTTCC
AX-89444977	**KUBO 179**	152169892	GAAGGTGACCAAGTTCATGCTGCCACAGTAAAAGTGATGACCTGA	GAAGGTCGGAGTCAACGGATTGCCACAGTAAAAGTGATGACCTGG	TATGCAAAGGGTCGACACTTCAAC
AX-89612309	**KUBO 180**	153574054	GAAGGTGACCAAGTTCATGCTAAACATCACTAGCCTTTATTCTCTCT	GAAGGTCGGAGTCAACGGATTAAACATCACTAGCCTTTATTCTCTCC	CAGGGACACTTTATAAGTACCCC
AX-89332583	**KUBO 181**	153916153	GAAGGTGACCAAGTTCATGCTCCACCTGTAAGTAGTAAGCACAAAAA	GAAGGTCGGAGTCAACGGATTCCACCTGTAAGTAGTAAGCACAAAAG	TTTTGCATTGACGGGTCCCAG
AX-89532109	**KUBO 182**	152849539	GAAGGTGACCAAGTTCATGCTGAGGCCCGTCGCGAGGCCGA	GAAGGTCGGAGTCAACGGATTGAGGCCCGTCGCGAGGCCGG	TGCTATGCAGCACTGGCGGCATAATA

The PCR protocol was used according to manufacturer instructions by LGC genomics. The following thermal protocol was used: 94°C 10 minutes – 10 cycles touchdown at 94°C 10 second and starting from 65°C to 57°C, 26 cycles at 94°C for 20 seconds and 57°C for 60 seconds.
Further re-cycling steps were applied in case a better separation of clusters was required using the following thermal protocol: 3 cycles at 94°C for 20 seconds and 57°C for 60 seconds. The reaction components are reported in Table 28 and Table 29 (Paragraph 5.2.4).

6.2.7 PCR specific assay

The PCR assay on the candidate gene *hox2* was composed of two primers forward and one primer reverse, specific to discriminate the presence/absence of the deletion in the genotype. The sequences of the three primers are the following:

Primer Fw2: 5'-CCACTAAAGATCATCCCCTGCTA-3'
Primer Rv1: 5'-CGGTCACGCAATATCCAC-3'
Primer Rv4: 5'-ACACTTATAGCTAACCATTCGGTG-3'

The PCR reaction was performed using GoTaq® G2 Flexi DNA Polymerase (Promega), using the following PCR components (Table 41)

Table 41: PCR components of hox2 specific assay using Promega Gotaq PCR kit.

Components	Working solution concentration (Ci)	Final Concentration (Cf)	Volume x 1-20µl in total (µl)
H_2O			4.54
PCR Buffer	5X	1X	4
MgCl2	25 mM	1.5 mM (highly stringent)	1.2
Primers F+R	10 µM $TE_{0.1}$ work (10µl Rv4+5µl Rv1+20µl Fw2+65 $TE_{0.1}$)	(normal) 0.4 µM = 400 nM of each	0.8
dNTPs	25 mM each	0.2 mM (fixed)	0.16
TAQ	1-5 Units / µl	0.5 – 1 Unit per reaction	0.1
DNA	10 ng / µl	100 ng per reaction	10

The PCR thermal protocol used for the reaction was the following, the fixed elongation time of 50 seconds was used to discriminate between presence or absence of *hox2* deletion (Table 42).

Table 42: thermal cycling protocol of the hox2 PCR specific assay.

TAQ activation	95°C for 5 min	1 cycle
denaturation	95°C for 45 sec	40 cycles
annealing	62°C for 45 sec	
elongation	72°C for 50 sec (1 min every kb)	
final extension	72°C for 5 min	1 cycle
hold	12°C for ∞	

6.2.8 Genetic interval evaluation

The gene interval included in the *GNI-2A* QTLs was explored using Biomart tool included in Ensembl plant database (Durinck et al., 2009) where high confidence genes were downloaded. The *hox2* expression pattern was extracted from the wheat expression browser (Borrill et al., 2016; Ramírez-González et al., 2018).

6.3 Results

6.3.1 QTL mapping on NCCR – background material

Starting from Unibo background material and work (Milner et al., 2016), Milner and colleagues used the NCCR mapping population and its genetic map to perform an association mapping analysis on yield traits and floret fertility architecture. As a results, a strong QTL on chromosome 2A (*QGns.ubo2A* QTL), hereafter referred as *GNI-2A*, was detected between genetic distances 84.6 cM and 87.6 cM, responsible for the regulation of the number of fertile florets per central spikelet (Table 43).

Table 43: Confidence interval detected for the GNI-2A QTL on the NCCR population using Illumina SNP 90K molecular markers. Chromosome and genetic distances on NCCR population are shown for each marker.

SNP code	CHR	cM
IWB27892	2A	84.59
IWB71483	2A	84.59
IWB33922	2A	84.59
IWA314	2A	84.59
IWB65847	2A	84.59
IWA581	2A	86.09
IWB72980	2A	87.59
IWB55871	2A	87.59
IWB43629	2A	87.59
IWB43630	2A	87.59
IWB74831	2A	87.59
IWB36149	2A	87.59
IWB72516	2A	87.59
IWB40453	2A	87.59
IWB12320	2A	87.59
IWB47511	2A	87.59
IWB24776	2A	87.59
IWB28709	2A	87.59
IWB66004	2A	87.59
IWA8491	2A	87.59
IWB2840	2A	87.59
IWB25154	2A	87.59
IWB43628	2A	87.59
IWB45265	2A	87.59

The most associated marker was IWA581, which has a phenotypic effect of 42% in the NCCR population and a -logPvalue of 40.36 (Table 44).
Based on the alignment on the Svevo pseudomolecules, the QTL interval detected was between 83 Mb and 155.9 Mb on chromosome 2A (Table 44).

Table 44: GNI-2A interval on NCCR population mapped on Svevo RefSeq v1.0 genome. The -logP-value based on Milner et al. (2016) is reported together with the r^2 effect (%).

SNP ID	-logPvalue NCCR population (Milner et al. 2016)	marker_R2 (%)	SVEVO_BLAST_physical (bp)
IWB26960	30.71	33.74	83,047,998
IWB20811	32.09	35.08	83,047,603
IWA4027	30.37	33.35	83,048,244
IWB34544	32.98	35.96	83,049,280
IWA3569	30.71	33.74	83,053,185
IWB329	32.32	35.45	84,319,271
IWB27190	32.98	35.96	84,319,311
IWB51634	32.50	35.44	84,630,155
IWB49088	31.24	34.22	85,683,530
IWB10760	32.32	35.45	86,161,856
IWB68419	31.63	34.58	86,216,726
IWB68420	32.30	35.09	86,216,748
IWB28709	33.34	36.20	86,685,873
IWB47511	33.61	36.61	86,703,545
IWB12320	33.61	36.61	86,711,316
IWB40453	34.29	37.13	87,135,104
IWB74831	34.27	37.10	87,155,112
IWB43629	34.44	37.16	87,157,898
IWB43630	34.44	37.16	87,157,964
IWB2840	33.49	36.24	87,167,532
IWB25154	33.00	35.73	87,897,800
IWB36149	34.44	37.16	88,276,287
IWB24776	32.58	35.61	88,642,793
IWB66004	33.49	36.24	88,646,196
IWB71483	34.80	37.48	88,921,233
IWB65847	34.10	36.96	89,571,125
IWA5240	36.09	38.68	93,029,271
IWB45265	33.61	36.61	87,132,698
IWB55871	34.29	37.13	87,134,639
IWB43628	33.00	35.73	87,157,500
IWB72980	33.77	36.58	87,188,075
IWB72516	33.75	36.65	87,889,127
IWA8491	33.49	36.24	88,276,987
IWA314	34.10	36.96	89,887,450
IWB75196	35.76	38.22	91,359,376
IWB71620	36.09	38.68	92,985,398
IWB71619	36.82	39.20	92,988,685
IWB23617	36.09	38.68	93,020,198
IWB3900	35.44	38.12	93,029,257
IWB59387	35.55	38.04	94,214,978
IWB23108	35.14	37.68	94,874,913
IWB48585	31.18	34.59	94,956,501
IWB48587	35.76	38.22	94,957,056
IWA1256	36.24	38.90	95,044,489
IWB11681	36.29	38.67	96,416,485

IWB73897	**36.09**	**38.68**	97,261,496
IWB664	**36.09**	**38.68**	97,261,596
IWB32206	**36.10**	**38.50**	97,278,301
IWA2245	**36.82**	**39.20**	98,205,852
IWB65613	**35.59**	**38.07**	99,459,539
IWA424	**36.29**	**38.67**	100,336,668
IWB33110	**33.60**	**39.24**	100,768,882
IWB72860	**36.82**	**39.20**	100,855,328
IWB62501	**36.35**	**39.36**	100,855,483
IWB72859	**36.82**	**39.20**	100,855,583
IWA7389	**35.57**	**38.14**	103,999,255
IWB49279	**34.95**	**37.34**	103,999,765
IWB5320	**35.57**	**38.14**	103,999,443
IWB5321	**35.55**	**38.04**	103,999,857
IWB27207	**35.59**	**38.25**	107,264,819
IWB65638	**36.09**	**38.68**	114,111,448
IWA6369	**36.82**	**39.20**	114,131,326
IWB45501	**36.82**	**39.20**	114,131,423
IWA5586	**35.55**	**38.04**	114,412,427
IWB45503	**36.29**	**38.67**	116,248,843
IWB45502	**36.82**	**39.20**	116,249,061
IWA1597	**36.09**	**38.68**	116,441,975
IWB21864	**36.90**	**39.46**	117,552,097
IWB34587	**36.82**	**39.20**	117,789,807
IWB37153	**36.82**	**39.20**	117,789,957
IWB71756	**36.09**	**38.68**	117,784,565
IWB66712	**34.29**	**37.12**	118,273,009
IWA994	**36.82**	**39.20**	120,418,376
IWB4905	**35.43**	**38.20**	126,515,332
IWB42663	**35.57**	**38.14**	126,517,238
IWB27678	**36.09**	**38.68**	126,517,436
IWB49366	**35.76**	**38.22**	129,682,673
IWB70278	**35.57**	**38.60**	137,504,107
IWB32396	**38.76**	**40.63**	145,172,046
IWB63013	**38.73**	**40.71**	145,174,846
IWB66099	**38.56**	**40.76**	145,693,567
IWA690	**39.59**	**41.70**	147,006,985
IWB32379	**38.73**	**40.71**	149,401,687
IWB65382	**38.84**	**41.17**	149,405,732
IWB32289	**38.27**	**40.61**	149,410,525
IWB32288	**39.87**	**41.83**	149,410,711
IWB24557	**39.87**	**41.83**	150,112,012
IWB1896	**39.69**	**41.78**	152,460,005
IWB69369	**36.48**	**38.83**	154,339,961
IWA581	**40.36**	**42.22**	154,732,761
IWB11613	**38.26**	**40.50**	155,892,451

6.3.2 KASP genotyping analysis

Starting from the most associated peak, -logP +/- 2 was considered to convert Illumina 90K SNP into KASP markers and narrow down the confidence interval, using marker probes from Illumina Infinium Array 90K SNP Chip.
Different KASP markers were developed within this interval, 6 proved to be polymorphic and were used to genotype the durum panel accessions from durum panel 1 (DP1) (Table 45) (Maccaferri et al., 2015). The genetic distance on the Svevo consensus map (Maccaferri et al., 2015) for each of the 6 KASP markers developed, ranged between 99 cM and 102 cM (different from the genetic distances previously reported by Milner et al., 2016) (Table 45).

Table 45: KASP markers developed from Illumina Infinium 90K SNP Chip inside the GNI-2A interval detected in NCCR population. The -logP, marker effect, position on Svevo RefSeq v1.0 assembly and genetic distance on Svevo consensus map are reported,

SNP_Consensus ID	KUBO ID	-log(p-value) NCCR population	marker_R^2 (%)	position_Svevo_bp	chr	cM-Consensu Map
IWB45503	**KUBO46**	36.29	38.67	116248843	2A	99
IWB45502	**KUBO45**	36.82	39.20	116249061	2A	99
IWB63013	**KUBO47**	38.73	40.71	145174846	2A	101.6
IWB24557	**KUBO44**	39.87	41.83	150112012	2A	102
IWB69369	**KUBO49**	36.48	38.83	154339961	2A	102
IWA581	**KUBO50**	40.36	42.22	154732761	2A	102

All primers proved to be polymorphic, codominant, and segregating for different haplotypes between contrasting durum parental lines Relief, Iride and the multi-parental line population founders Neodur, Colosseo, Claudio and Rascon/2*Tarro. The following KASP markers, KUBO 45, KUBO 46, KUBO 47, KUBO 44, KUBO 49, KUBO 50 ranged between 116 Mb and 155 Mb on chr2A of *Triticum turgidum* cv Svevo pseudomolecule, and were used to genotype the DP1, a panel composed by a subselection of 167 genotypes out of 300 (Maccaferri et al., 2015) (Table 46).

Table 46: Summary of genotyping results for KASP markers on DP1 using KUBO45, KUBO 46, KUBO 47, KUBO 44, KUBO 49, KUBO 50.

DP1 population	n° of lines
Total number of lines	167
WT haplotype	158
Mutant haplotype	4
Recombinant lines	5
% mutant lines	2.40
% wt lines	94.61
% recombinant lines	2.99

The amplification plots show the segregation between *GNI-2A haplotype*, wt genotypes and artificial heterozygotes in the durum panel accessions (Figure 66). All the markers tested show a good discrimination capacity and codominance between opposed parental lines.

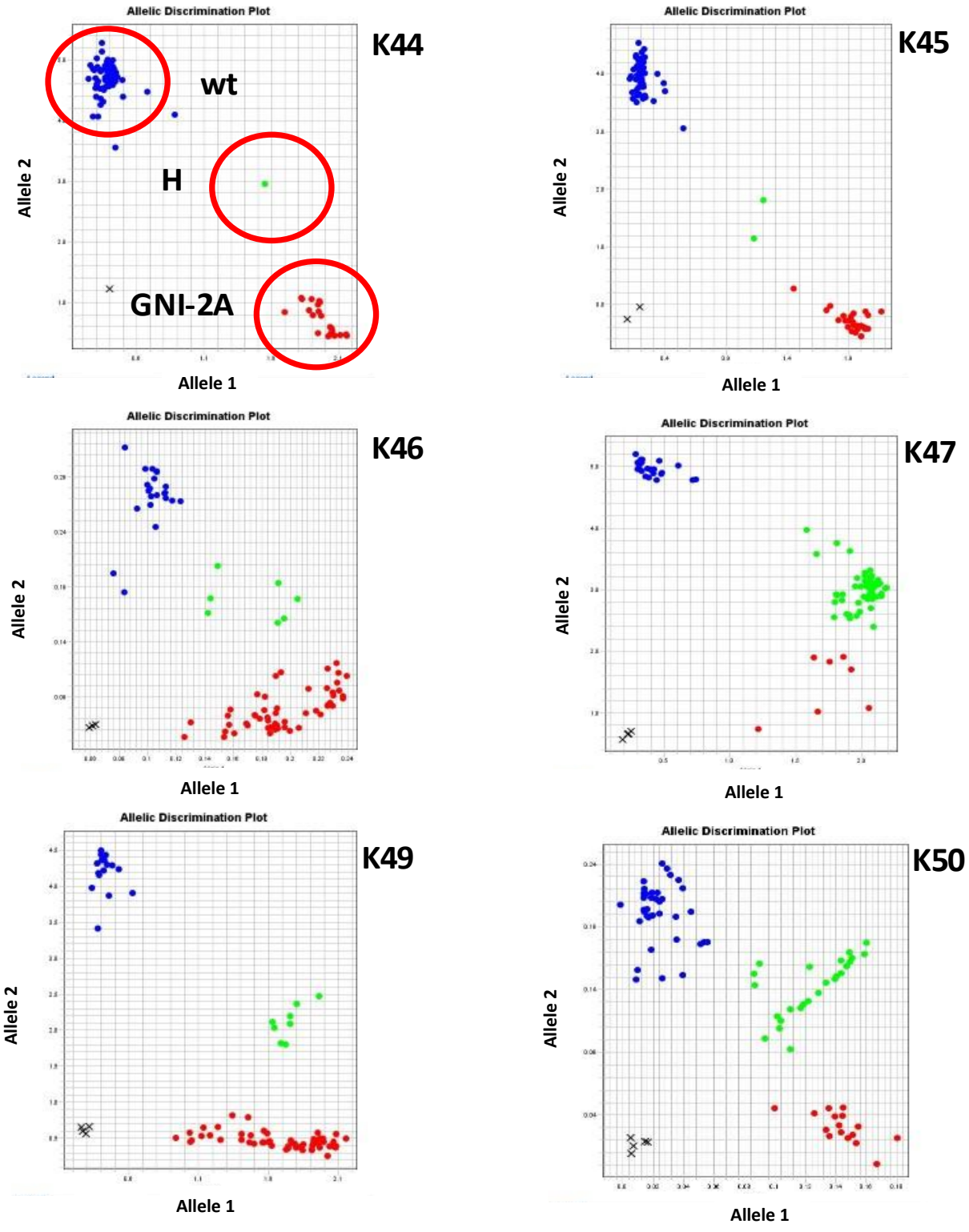


Figure 66: Amplification plots for KUBO 44. KUBO 45, KUBO 46, KUBO 47, KUBO 49 and KUBO 50 on DP1. Wt and GNI-2A alleles are highlighted in blue and red respectively, the heterozygotes are in green.

The fertility haplotype was detected in less than 2.4% of analysed elite durum panel genotypes, resulting as a rare allele (minor than 5%) in modern elite variety panel (Table 46). To confirm these results, the allele frequency for each SNP 90K marker converted into KASP was calculated in the DP1 genotyping array, strengthening the fact that single markers and the *GNI-2A* haplotype corresponds to rare alleles in the population responsible for an increase in spike fertility trait.

Marker ID	SVEVO_BLAST_physical	KASP	CHR	Altar_84	RASCON/2* TARRO	Iride	Saragolla	Claudio	Neodur	Colosseo	Svevo
IWB45503	116,248,843	KUBO_46	2A	T	T	T	T	G	G	G	G
IWB45502	116,249,061	KUBO_45	2A	G	G	G	G	T	T	T	T
IWB63013	145,174,846	KUBO_47	2A	C	C	C	C	A	A	A	A
IWB24557	150,112,012	KUBO_44	2A	G	G	G	G	T	T	T	T
IWB69369	154,339,961	KUBO_49	2A	A	A	A	A	C	C	C	C
IWA581	154,732,761	KUBO_50	2A	C	C	C	C	T	T	T	T

Figure 67: NCCR founders and GNI-2A haplotypes based on KASP markers converted from 90K SNP Chip.

As shown in Figure 67, the rare haplotype corresponds to cultivars Altar-84, Rascon/2*Tarro, Iride and Saragolla, which have a *GNI-2A* haplotype and show increased number of florets per spikelet. This is in contrast with NCCR parental lines Neodur, Claudio, Colosseo and the genome sequenced reference cultivar Svevo, which have a different haplotype and do not have the increase in spike fertility.

As a confirmation of the haplotypes detected on modern cultivars from Illumina Infinium 90K SNP Chip, resulting KASPs from the KASP genotyping on DP1 show that the elite cultivars with the *GNI-2A* haplotype are: Iride (which carries the Altar-84 haplotype), Saragolla, Sculpture and Cuspide, which have an increase in spike fertility (Figure 68).

Durum cultivars		K46	K45	K47	K44	K49	K50
Acadur	Italy	A	A	B	B	H	H
Achille	Italy	B	B	B	B	X	H
Adone	Italy	X	B	B	B	B	B
Alemanno	Italy	X	B	B	B	B	B
Amina	ICARDA	B	B	B	B	B	B
Amria	ICARDA	B	B	B	B	B	B
Anvergur	France	B	B	B	B	H	B
Augusto	Italy	B	A	B	B	H	B
Aureo	Italy	B	B	X	B	H	B
azeghar_2	ICARDA	B	B	B	B	B	B
Babylone	France	B	B	B	B	H	B
Bacardi	Italy	B	B	B	B	B	B
Casteldoux	France	B	B	H	B	B	B
Ceedur	France	A	H	B	B	B	B
Cesare	Italy	B	B	B	B	H	B
Cham-1	ICARDA	B	B	B	B	B	B
Cincinnato	Italy	B	B	B	B	B	B
Claudio	Italy	B	B	B	B	B	B
Claudio Sis	Italy	B	B	B	B	B	B
Cndo/Primadur	Italy	B	B	X	A	A	A
Colombo	Italy	B	B	B	B	H	B
Corallo	Italy	B	B	B	B	B	B
Core	Italy	B	B	B	B	B	B
Credit	Italy	B	B	A	A	A	A
Cuspide	Italy	A	A	A	A	A	A
DGE1	USA	X	B	X	B	B	A
desert king	USA/desert durum	B	B	B	B	B	B
Diamante	Italy	B	B	B	B	B	B
Duroi	Italy	B	B	B	B	X	B
Dylan	Italy	B	B	B	B	B	B
Edmore	Italy	B	B	B	B	B	B
Emilio Lepido	Italy	B	B	B	B	B	B
ettore	Italy	B	B	B	B	H	B
FD14DW021	France	B	B	H	B	B	B
Flavio	Italy	B	B	B	B	B	B
Furio Camillo	Italy	B	B	B	B	H	B
Gibraltar	France	B	B	B	B	B	B
Gidara	ICARDA	B	B	H	B	B	B
Ginseng	Italy	B	B	B	B	H	B
Giulio	Italy	B	B	B	B	B	B
Grazia	Italy	B	B	B	B	B	B
Haby	ICARDA	B	B	B	B	B	B
ica mor.	ICARDA	B	B	B	B	B	B
Iride*** (TaHOX2 allele from Altar84-Gallareta)	Italy	A	A	A	A	A	A
Joyau	France	B	B	B	B	B	B
Kanakis	Italy	B	B	B	B	B	H
karim	CIMMYT	B	B	B	B	B	B
Karur	France	B	B	B	B	B	H
Kiko nick	Italy	B	B	B	B	B	H
Kronos	USA/desert durum	B	B	B	B	B	B
Kyle	USA	B	B	B	B	B	B
Lambro	Italy	H	B	B	B	B	B
Langdon	USA	B	B	B	B	B	B
Levante	Italy	B	B	B	B	B	B
liberdur	France	B	B	B	B	B	B

Durum cultivars		K46	K45	K47	K44	K49	K50
Lloyd	USA/France	B	B	B	B	B	B
M1084	ICARDA	B	B	B	B	B	B
M20	ICARDA	B	B	B	B	B	B
Maestà	Italy	B	B	B	B	B	B
Maier	USA	B	B	B	B	B	B
Marakas	Italy	A	A	B	B	B	B
Marco Aurelio	Italy	B	B	B	B	B	H
Mario	Italy	B	B	B	B	B	X
Massimo Meridio	Italy	B	B	B	B	B	B
Miki	ICARDA	B	B	B	B	B	B
Mindum	USA	B	B	B	B	B	B
Miradoux	France	B	B	B	B	B	B
Morse	USA	B	B	B	B	B	B
Nadif	Italy	B	B	B	B	B	B
Natal	Italy	B	B	B	B	B	B
Neodur	France	B	B	B	B	B	B
Nobilis	France	B	B	B	X	B	H
Obelix /	Italy	B	B	B	B	B	B
Odisseo	Italy	B	B	B	B	B	B
Ofanto	Italy	B	B	B	B	B	B
omrabi_5	ICARDA	B	B	B	B	B	B
Opera	Italy	B	A	B	B	B	H
Orizzonte	Italy	B	B	B	B	B	H
Orobel	France	B	B	B	B	B	B
Ovidio	Italy	B	B	B	B	B	H
Pellisier	Tunisia	B	B	B	B	B	B
Pescadou	France	B	B	B	B	B	B
Pi Greco	Italy	B	B	B	B	B	B
plata_10/6/mqe/	CIMMYT	A	A	B	B	B	B
PR22D84	Italy	B	B	B	B	B	B
Produra	CIMMYT	B	B	B	B	B	B
Provenzal	France	B	B	H	H	H	H
Ramirez	Italy	H	H	B	B	B	B
Relief	France	B	B	B	B	B	B
Sachem	France	B	B	B	B	B	B
san carlo	Italy	B	B	B	B	B	B
Saragolla	Italy	A	A	A	A	A	A
sculpture	France	A	A	A	A	A	A
Secolo	Italy	B	B	B	B	B	X
Sy Cisco	Italy	X	H	B	B	B	H
Sy Lido	Italy	B	B	B	B	B	H
teodorico	Italy	B	B	B	B	B	H
Textur	France	B	B	B	B	B	B
tirex	Italy	B	B	B	B	B	B
Tito Flavio	Italy	B	B	B	B	B	H
Tiziana	Italy	B	B	B	B	B	B
Trinakria	Italy	B	B	B	B	B	B
Valnova	Italy	B	B	B	B	B	B
Varano	Italy	X	B	B	B	X	B
Vatan	East Europe	B	B	B	B	B	B
Vendetta	Italy	B	B	B	B	B	B
West Bred Turbo	USA/ desert durum	B	B	B	B	B	B
zaghar.	ICARDA	B	B	B	B	B	B
zardak	IRAN	B	B	B	B	B	B
zeina_1	ICARDA	B	B	B	B	B	B

Figure 68: KASP genotyping analysis on 167 genotypes from DP1 using KUBO 44. KUBO 45, KUBO 46, KUBO 47, KUBO 49 and KUBO 50. A stands for GNI-2A allele and B for wt allele, H stands for heterozygotes

6.3.3 Phenotypic analysis

The number of fertile florets per central spikelets and other spike architecture parameters were scored in the Relief x Iride population during 2019 and 2020 growing season. The number of fertile florets per central spikelet was monitored by Milner et al (2016) also in the four-way NCCR population (background material). As visible in Figure 69, the spike morphology is differentiated between the parental lines of the NCCR population. The Rascon / 2*Tarro spike show higher number of spikelets with a denser spike in comparison to the other parental lines (Figure 69).

Figure 69: spike differences between the four parental founders of NCCR population.

Milner et al (2016) compared the number of fertile florets per central spikelet at different growing stages following BBCH wheat growth scale (Zadoks et al., 1974), from GS49 to GS70, across the 4 NCCR parental lines and Altar-84 variety (CYMMIT line that originated the Rascon /2*Tarro haplotype). Figure 70 shows that Rascon /2*Tarro and Altar-84 have a higher number of fertile florets per central spikelet across the different growing stages, reaching the maximum values (9) between GS65 and GS69.

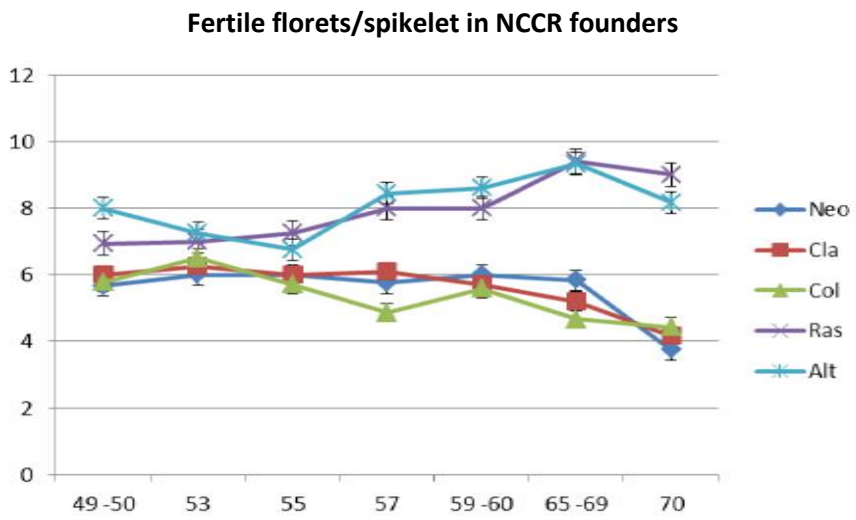


Figure 70: different number of fertile florets per central spikelets across the NCCR parental lines in the different growing stages, from BBCH 49-50 to BBCH 70.

As regards the Relief x Iride biparental population, phenotypic distribution of number of florets per central spikelet was assessed by data analysis during 2019 and 2020 growing season after harvest. The distribution of the dataset was bimodal, showing peaks on the average number of florets which segregates for the parental average phenotypes, namely 4.5 and 6.3 florets per central spikelet (Figure 71).

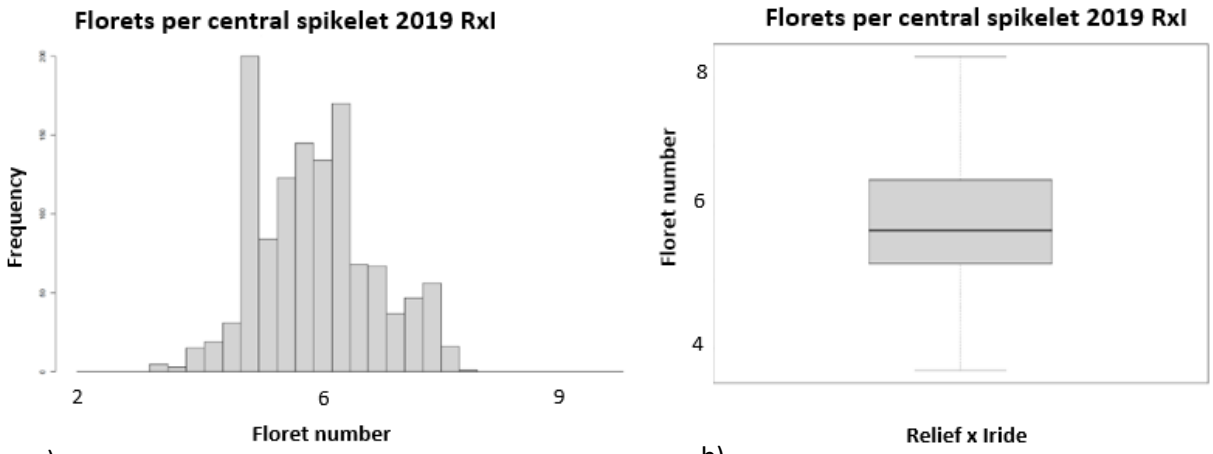


Figure 71: phenotypic distribiution of Relief x Iride population in 2019. a) the distribution is bimodal, b) the first and fourth quartile range goes from 3 to 8.5 fertile florets per central spikelets

The range of phenotypic diversity observed in the population went from 3 to 8.5 florets per central spikelet, broad sense heritability reported was 88% (Table 47).

Table 47: Descriptive statistics of Relief x Iride population from 2019, and heritability value.

Descriptive statistics 2019	Florets per central spikelet
min	3
max	8.5
range	5.5
median	5
mean	4.98
SE.mean	0.02
σ^2	1.08
ST. DEV	1.04
CV	0.21
h^2	0.88

During the 2019/2020 growing season, Relief x Iride was phenotyped for spike fertility trait with the same methodology used for the previous season, in two replicates and 6 selected spikes per genotype.
As previously showed for the 2018/2019 growing season, the number of florets per central spikelet distribution detected was bimodal with the following peak averages: 3.8 for wt genotypes and 4.9 for *GNI-2A* genotypes (Figure 72). The average values were a little bit lower than expected, especially based on 2019 data, however presence of diseases and differential growing seasons could hamper the floret development.

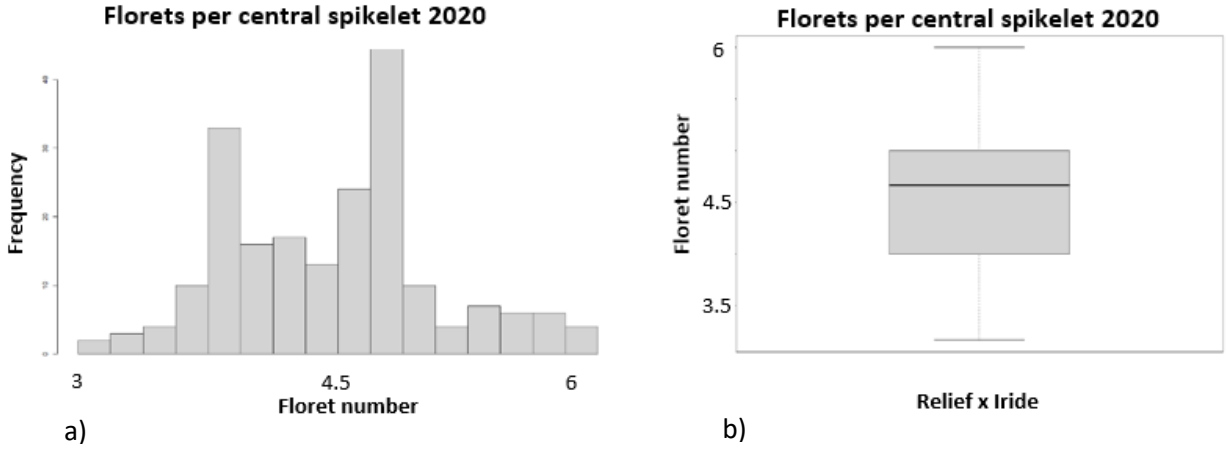


Figure 72: phenotypic distribiution of Relief x Iride population in 2020. a) the distribution is bimodal, b) the first and fourth quartile range goes from 3.5 to 6 fertile florets per central spikelets.

The range was between 3 and 6 florets per central spikelet and the heritability calculated value was 89% (Table 48).

Table 48: Descriptive statistics of Relief x Iride population from 2020, and heritability value.

Descriptive statistics 2020	Florets per central spikelet
min	3
max	6
range	3
median	5
mean	4.57
SE.mean	0.02
CI.mean.0.95	0.04
σ^2	0.59
ST.DEV	0.77
CV	0.17
h^2	0.89

Based on the analysis of 2019 and 2020 for Relief x Iride population, the phenotypic distribution shows that the fertile florets per central spikelets segregates within the population for the multi-year trials. This confirms the presence of a strong QTL (*GNI-2A*) which regulates the trait with a strong genetic component between the different generations, as the heritability value are close to 90%.

6.3.4 QTL fine mapping

The *GNI-2A* interval was explored considering the chromosomic region with the highest association scores and effect on the NCCR population, considering -logP +/-2 within the region, thus evaluating the interval between 150 Mb and 154.9 Mb. The considered interval was included between KUBO 44 and KUBO 49 (Figure 73).

The confidence interval was confirmed by KASP PCR assay (using KUBO 44 and KUBO 49) performed on Relief x Iride F_6 population, extracting DNA from selected spikes harvested during 2019/2020 growing season.

As a result, the analysis showed KASP primers segregation inside the Relief x Iride genotypes, accordingly to Rascon/2*Tarro (*GNI-2A* haplotype) and Svevo (wt haplotype).

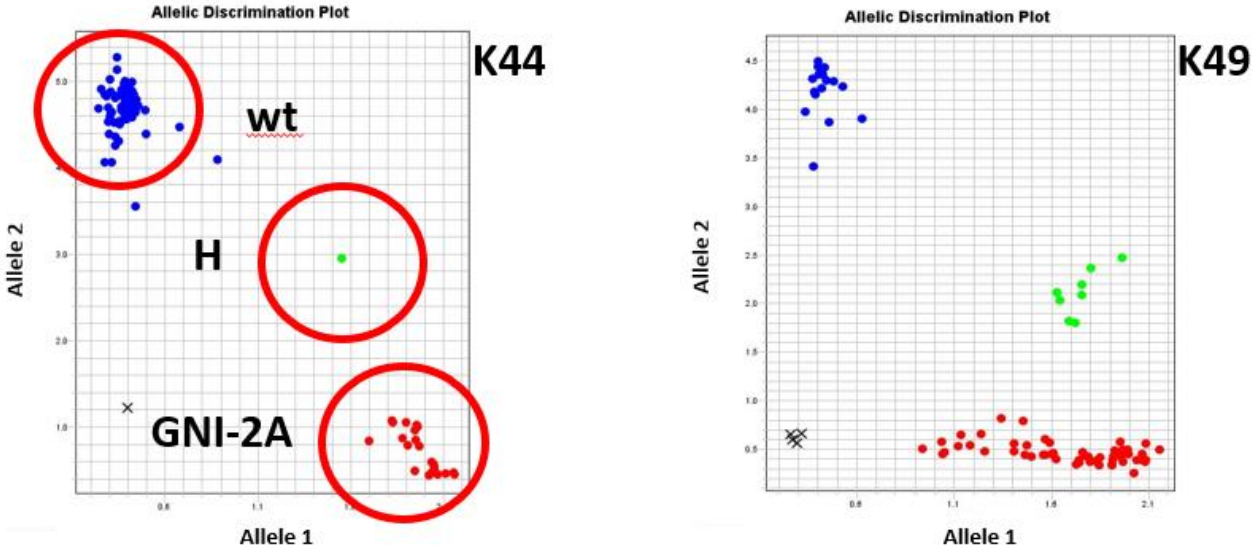


Figure 73: Amplification plots of KUBO 44 and KUBO 49 between GNI-2A and wt genotypes on Relief x Iride population from 2020.

The identified interval between KUBO 44 and KUBO 49 was approximately of 4.9 Mb, much shorter than the interval detected in NCCR population (Table 44). As a consequence of the Relief x Iride F_6 genotyping, 401 genotypes showed the wt haplotype, 326 accessions had the *GNI-2A* haplotype, and 10 genotypes were recombinant between KUBO 44 and KUBO 49. The linear model between haplotype and number of florets per central spikelet was computed for 2020 Relief x Iride dataset, where the detected r^2 was equal to 0.46. A significant difference was detected based on ANOVA performed on the two genotypic groups (*GNI-2A* and wt) and the number of florets per central spikelet (Figure 74).

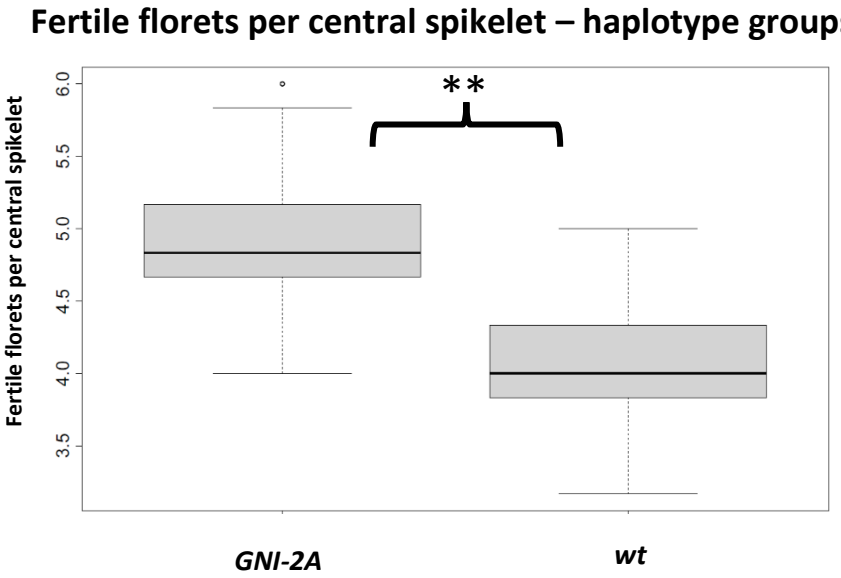


*Figure 74: boxplots between GNI-2A and wt genotypes in Relief x Iride population. The difference between the boxplot is significant (** = <0.05). The wt and GNI-2A medians are approximately 4 and 4.7 respectively.*

In order to further fine map the confidence interval, the Affymetrix SNP Chip 420K was exploited. The SNP array was aligned on 55 durum elite varieties (background material Unibo) and polymorphic SNP between wt and *GNI-2A* varieties haplotypes were chosen to be converted in new KASP primers. Eight KASP markers proved to be polymorphic between population parents, Relief and Iride, by KASP PCR analysis on 10 recombinants detected between KUBO 44 and KUBO 49.
In order to strengthen the phenotypic data of recombinant genotypes, 3 central spikelets were phenotyped for number of florets for each genotype in two replicates, the final phenotypic average data was obtained.
Based on the phenotypic data and fined mapped genotypic data, the interval was narrowed down between 150.35 Mb and 153.92 Mb on chr2A (3.9 Mb), critical recombination event were detected between KUBO 44 / KUBO 177 and KUBO 49 / KUBO 181 as Relief x Iride recombinant lines 1140, 1170 and 1172 showed an increase number of fertile florets per spikelet ranging from 4.47 to 5.39 (Table 49).

Table 49: Fine mapped interval on Svevo RefSeq v1.0 genome on chromosome 2A. The wt or GNI-2A haplotypes are reported, based on the phenotype (number of florets per central spikelet) the interval was fine mapped between KUBO181 and KUBO177

	Svevo bp	572	651	1011	1125	1140	1170	1172	1319	1405	801
k47	145174846	wt	GNI-2A	wt	GNI-2A	GNI-2A	wt	GNI-2A	wt	wt	wt
k44	150112012	wt	GNI-2A	wt	GNI-2A	GNI-2A	wt	GNI-2A	wt	wt	wt
k177	150351083	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k178	151565662	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k179	152169892	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k169	152460005	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k182	152849539	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k173	153155184	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k175	153760282	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k181	153916153	wt	wt	wt	wt	GNI-2A	GNI-2A	GNI-2A	wt	wt	wt
k49	154339961	GNI-2A	wt	GNI-2A	wt	wt	GNI-2A	wt	GNI-2A	GNI-2A	wt
k50	154732761	GNI-2A	wt	GNI-2A	wt	wt	GNI-2A	wt	GNI-2A	GNI-2A	wt
Florets/central spikelet 2020		3.47	3.75	3.97	3.83	4.47	4.54	5.39	4.13	4.33	4.06

6.3.5 *GNI-2A* genetic interval exploration

The genetic interval was explored in Ensembl Plants database, exploring the region between KUBO 44 and KUBO 49 on chromosome 2A of *Triticum turgidum* cv. Svevo. The fine mapped interval contained 8 High confidence gene and 13 low confidence gene (Table 50).

Table 50: GNI-2A gene interval between KASP markers KUBO44 and KUBO49. Gene positions and functions are reported, the candidate gene is highlighted in green.

Gene_stable.ID	Gene_start	Gene_end	Confidence	Human_Readable_Description
k44	150112012			
TRITD2Av1G065720	150154760	150155062	LC	LisH/CRA/RING-U-box domains-containing protein
TRITD2Av1G065760	150277260	150277673	LC	Na(+)/H(+) antiporter NhaB
TRITD2Av1G065790	150351030	150351401	LC	NAD kinase 2
k177	150351083			
TRITD2Av1G065810	150406349	150407143	LC	NAC domain-containing protein 21/22
TRITD2Av1G065820	150418530	150419438	LC	Succinate dehydrogenase assembly factor 2, mitochondrial
TRITD2Av1G065930	151044680	151045171	HC	Protein kinase superfamily protein
TRITD2Av1G065990	151156448	151157290	LC	ATP-dependent RNA helicase dhx8
TRITD2Av1G066020	151336034	151336531	LC	ARM repeat superfamily protein
TRITD2Av1G066040	151392407	151392739	LC	nuclear RNA polymerase C2
TRITD2Av1G066050	151476083	151492870	HC	Homeobox leucine zipper protein
k178	151565662			
TRITD2Av1G066070	151574182	151574706	LC	RING/U-box superfamily protein
k179	152169892			
TRITD2Av1G066110	152458187	152459853	HC	Polyol transporter
k169	152460005			
TRITD2Av1G066120	152477625	152479031	HC	Polyol transporter
TRITD2Av1G066140	152700098	152700412	LC	Subtilisin-like protease
TRITD2Av1G066160	152785613	152786326	LC	Dual 3',5'-cyclic-AMP and -GMP phosphodiesterase 11A
TRITD2Av1G066170	152786434	152787132	LC	Iron-sulfur cluster insertion protein ErpA 1
TRITD2Av1G066180	152814040	152815294	LC	Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase
TRITD2Av1G066190	152819657	152819863	LC	Histidine biosynthesis bifunctional protein HisB
TRITD2Av1G066220	152844465	152846986	LC	CwfJ-like family protein / zinc finger (CCCH-type) family protein
TRITD2Av1G066230	152849223	153072120	HC	Polyol transporter
k182	152849539			
k173	153155184			
TRITD2Av1G066240	153157347	153158896	HC	Sorbitol transporter
TRITD2Av1G066280	153463708	153464085	LC	NAD(P)-binding Rossmann-fold superfamily protein
TRITD2Av1G066300	153567357	153571215	HC	Mitochondrial transcription termination factor family protein
TRITD2Av1G066310	153573105	153575601	LC	Zinc finger C-x8-C-x5-C-x3-H type family protein
TRITD2Av1G066350	153723067	153723638	HC	Sodium channel protein type 5 subunit alpha
TRITD2Av1G066360	153727715	153729013	HC	F-box protein
TRITD2Av1G066380	153751787	153752158	LC	nucleoporin
TRITD2Av1G066390	153756304	153761403	HC	Proteasome subunit alpha type
k175	153760282			
TRITD2Av1G066460	153915102	153924089	HC	1-phosphatidylinositol-3-phosphate 5-kinase
k181	153916153			
TRITD2Av1G066500	154060075	154061397	LC	zinc knuckle (CCHC-type) family protein
TRITD2Av1G066520	154179985	154180290	LC	DNAse I-like superfamily protein
TRITD2Av1G066560	154338005	154339669	HC	GRAS transcription factor
k49	154339961			

Based on background analysis performed by a collaboration between Unibo and IPK (Milner et al., 2016), a probable candidate gene was identified in *TRITD2Av1G066050*, homeobox leucin zipper transcription factor (Figure 75), which corresponds to the *hox2* gene, identified as an orthologue of *vrs1* (Sakuma et al., 2017) and a paralogue of the *GNI-1* in *Triticum aestivum* (Sakuma et al., 2019).

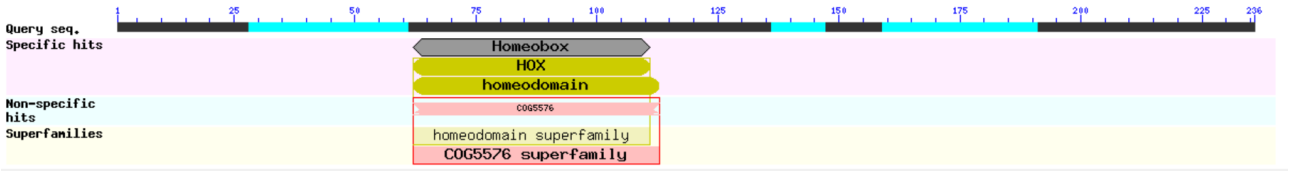


Figure 75: functional domain of hox2 gene, homeobox leucin zipper domain from NCBI blast database.

By background information from the collaboration between Unibo and IPK, a resequencing analysis was performed on different durum cultivars, some of which share the Altar-84 haplotype (*GNI-2A*), showing that the gene has a 4 kbp deletion in in Rascon/2*Tarro and Altar-84 related germplasm. This mutation seemed to be strongly correlated with spike fertility trait, strengthening the fact that the *hox2* gene seem to be a promising candidate for the increased number of fertile florets per central spikelet. The presence of the deletion in the candidate gene was confirmed by a PCR specific assay, where three primers are used (Figure 76). Based on the presence or absence of the deletion, using the same elongation time for the PCR reaction (50 seconds), it is amplified a fragment of 0.2 kbp (in the absence of the deletion, primers Fw2 and Rv1 produce the PCR product in wt genotype) or a fragment of 0.9 kbp (presence of the deletion, *GNI-2A* genotype).

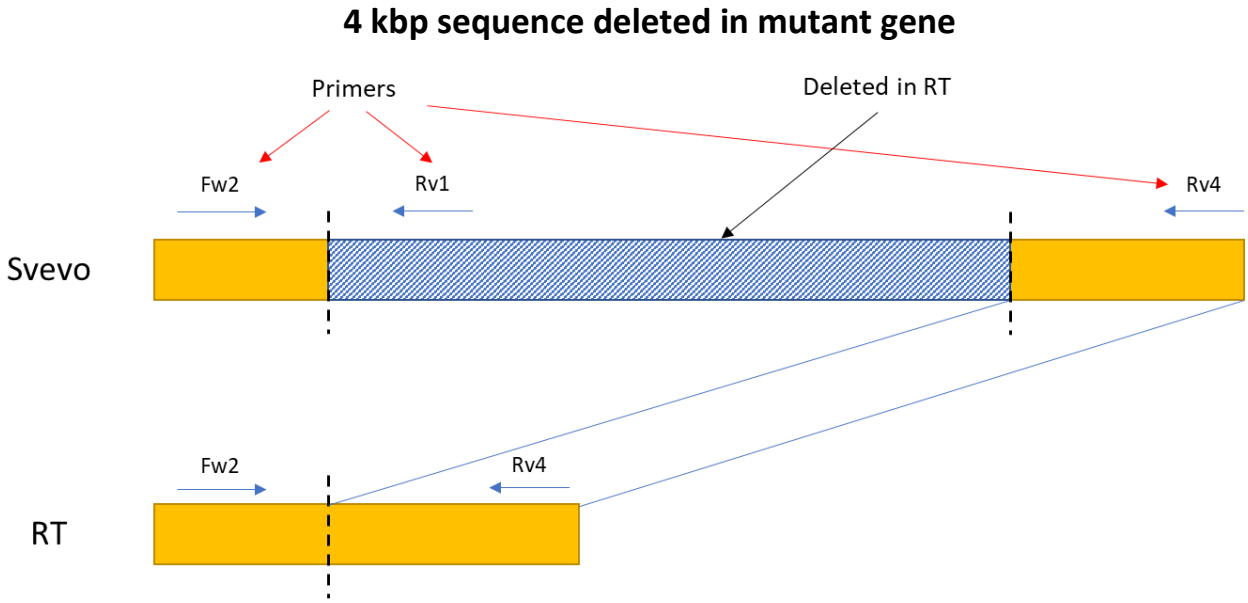


Figure 76: difference between deleted (RT) and wt sequence (Svevo). The blue rectangle corresponds to the 4kbp deletion present in hox2 mutant gene. The primers that produce the PCR fragment in the absence of the deletion are Fw2 and Rv1, on the other hand, when there is the deletion, primers Fw2 and Rv4 produce a 0.9 PCR fragment.

The PCR results on parental Rascon/2*Tarro, Svevo, and artificial heterozygote (H) showed that the PCR system is efficient and confirmed the presence of the deletion in Rascon/2*Tarro haplotype, and that both alleles are present in the artificial heterozygote (H) performed by merging equal volumes of Rascon/2*Tarro and Svevo (Figure 77).

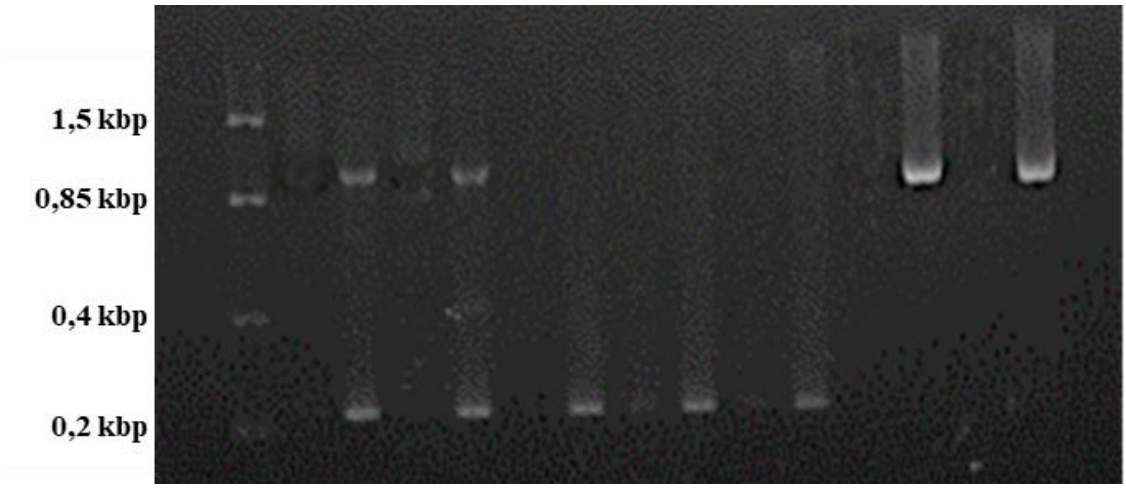


*Figure 77: PCR result of PCR specific assay on hox2 gene between contrasting GNI-2A varieties Svevo (Sv) and Rascon/2*Tarro (RT). The artificial heterozygote was composed by equal volumes of Sv and RT.*

The *hox2* expression was explored using Wheat expression browser database (Ramírez-González et al., 2018) in different tissues and at different developmental stages of the *Azurnhaya* cultivar (Figure 78).

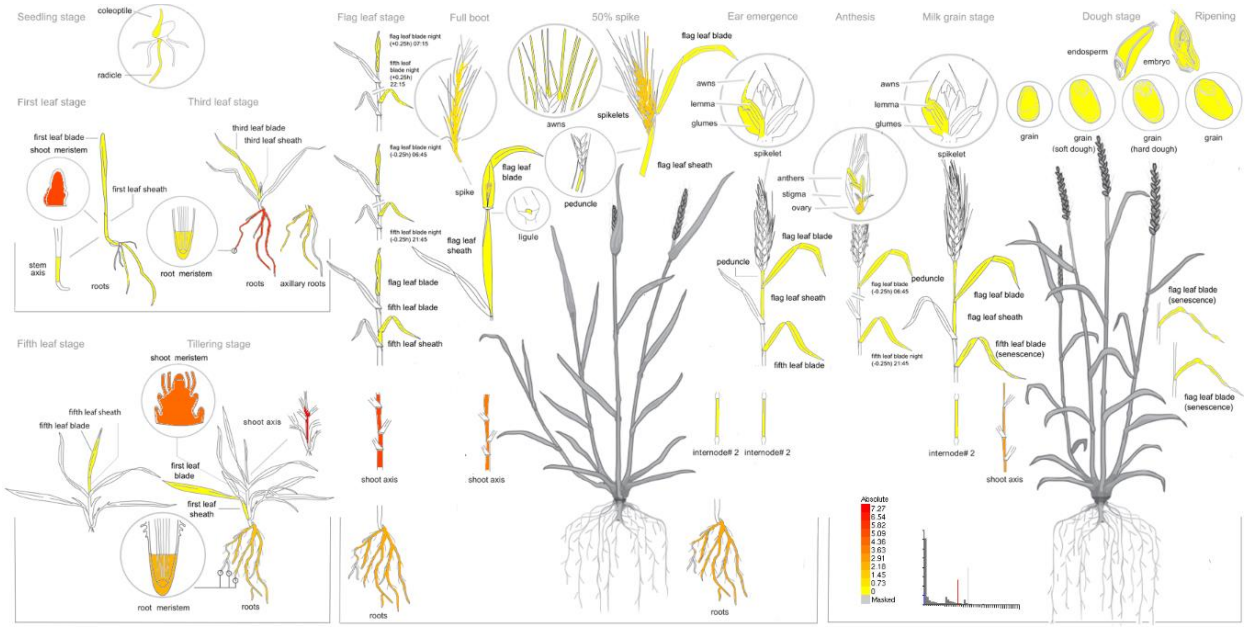


Figure 78: hox2 gene expression in wheat expression browser database. The level of expression in different tissues at different stages of Azurnhaya cultivare are shown in tpm expression unit.

The gene expression is very high in early developmental stages, during spike meristem differentiation. At adult stages the expression decreases but remains stable especially in mature spike.

6.4 Discussion

The reported work describes the characterization of the QTL *GNI-2A* in the mapping population Relief x Iride F_6 and in the durum panel 1 (Maccaferri et al., 2015). The QTL was initially mapped in the 4-way durum cross population NCCR on chromosome 2A, leading to an increase of 0.55 grains per spikelet (Milner et al., 2015). The *GNI-2A* haplotype derives from the CYMMIT line Altar-84 which is related to the NCCR parental lines Rascon/2*Tarro. The herein work report a complete phenotypic and genotypic characterization of the biparental population Relief x Iride F_6, Iride has the same haplotype and is strongly related with the Altar-84 CYMMIT line. Iride has an increase in the number of florets per central spikelet, causing an increase in spike fertility and yield but leading to the abortion of apical florets during spikelet differentiation. Multi-years field trials and spike fertility phenotyping showed a segregation of the *GNI-2A QTL* in the Relief x Iride population, showing a bimodal distribution between all the genotypes. As both Iride and Rascon /2*Tarro derived from the CYMMIT line Altar-84, the range of values detected for Relief x Iride was similar to the ones described in NCCR multi-parental cross (Milner et al., 2016), with a stronger bimodal distribution due to the high QTL segregation in the biparental population. The fine mapping occurred converting 11 KASP markers from Illumina Infinium SNP 90K Chip and Affimetrix Axiom 420K narrowing down the interval in a 3.9 Mb. The fine mapped interval contains about 15 high confidence (HC) genes, but a strong candidate corresponds to the *hox2* genes, that has a 4 kbp deletion in Altar-84 related haplotypes. This gene is a strong candidate as its molecular function belongs to a gene family strongly involved in spike development and floret differentiation. In fact, this gene codifies for a homeobox leucin zipper transcription factor, which is the same protein domain detected for other genes involved in spike fertility and architecture morphology.
For example, the *vrs* genes (*vrs1-5*) are deeply involved in barley spike morphology with multi-rows spike leading to an increased number of grains per spike, in particular *vrs-1* codifies for a HD-Zip1 protein domain typical of transcription factors protein (Komatsuda et al., 2007; Sakuma et al., 2013). In addition, the identified *hox2* gene is a paralogue of *GNI-1* which codifies for a similar kind of protein (Sakuma et al., 2019). Paralogue genes correspond to multiple copies of the same sequence in the genome. The *GNI-1* gene is on the same chromosome (chr2A) of *hox2* on *GNI-2A*, approximately 400 Mb apart. *GNI-1* has a single amino acidic mutation (N105Y) that prevents from the inhibition of apical florets abortion (Sakuma et al., 2019). This function can be connected with the *GNI-2A* mutation on Altar-84 related genotypes, as the genes carrying the 4 kbp deletion lead to an increase in spike fertility preventing apical florets abortion. In fact, it has been demonstrated that the overexpression of the *hox2* gene produces plants with reduced fertility (Wang et al., 2017). Thus, both *hox2* and *GNI-1* wild type genes regulate the apical floret abortion reducing the number of florets per spikelet. On the other hand, the mutation for both genes that causes a loss of functions is responsible for an increase of grain number per spikelet where apical florets are not aborted.
It is evident that there is a strong connection between *GNI-2A* and *GNI-1* regarding the phenotypic effect, the function on the development of florets per spikelet and the tissues/developmental stages of expression. Furthermore, the segregant population Relief x Iride could also be influenced by the action of *GNI-1*, which could not be excluded from the phenotypic expression. However, based on the 20 KASP markers used to genotype the population and on the phenotypic analysis, it can be

concluded that the *GNI-2A* QTL has the major effect on the spike fertility trait in the segregant Relief x Iride population.
The KASP mapping on the durum panel germplasm revealed that the *GNI-2A* QTL is a rare allele in the modern breeding varieties. Rare alleles are defined with a frequency lower than 5% in the population. The fact that the *GNI-2A* haplotypes is rare in the durum panel probably means that the mutation, occurred in the Altar-84 variety and related genotypes based on pedigree, has occurred recently in CYMMIT germplasm and it is not spread in the modern varieties of durum panel 1.
It is worth pointing out that the flowering time regulation and spike fertility are regulated by different type of genes belonging to similar protein family, such as transcription factors with DNA binding domains. For example, similar functions and protein domains are shared between *vrs* genes and *GNI* in barley and wheat (Komatsuda et al., 2007; Sakuma et al., 2013; Sakuma et al ., 2019). For example, other genes involved in the spike morphology traits are the *Q* gene which encodes for a AP2-like transcription factor (Simons et al., 2006). The mutation of this gene is responsible of an increase rachilla length with more than 12 flowers, but almost completely sterile.
As regards the location and level of transcription of *hox2* gene, higher expression was detected on *Triticum aestivum* from the wheat expression database browser between first leaf and tillering stage, in immature spikes tissues (Figure 78). The expression pattern is similar to the one reported for *GNI-1*, whose expression was increased up the white anther/green anther stage (based on Kirby scale) and then it is reduced in the following step of spike developmental stage (Kirby, 1974; Sakuma et al., 2019). Both genes seem to mainly work on the immature spike developmental stages, regulating spikelet formation and apical floret development. The transcriptional level is then lowered at later stages of the spike development for both genes, meaning that floret development was established at early spike development stages.
To conclude, the herein work reports the characterization of spike fertility trait on the biparental population Relief x Iride, exploiting QTL mapping analysis performed on NCCR population (Milner et al., 2016). The work identifies a bimodal segregation of *GNI-2A* QTL in the biparental population, which strongly influences the floret phenotype. The strong candidate gene for the *GNI-2A* QTL is represented by *hox2*, which has a 4 kbp deletion in the mutant genotypes which is rare by KASP analysis in modern durum panel meaning that it represents a recent event occurred in Altar-84 and related CYMMIT genotypes. The herein work developed KASP markers, polymorphic and high informative for the *GNI-2A* QTL, that can have a practical outcome for marker assisted selection and in breeding programs to select for increase fertility haplotypes.

7. Conclusions and future perspectives

The herein reported PhD thesis was developed in the framework of Innovar H2020 project. This project aims at using genomics, transcriptomics, high throughput phenotyping techniques and machine learning to update and improve varietal characterization and registration procedure used in Europe for VCU and DUS protocols. The focus was on durum wheat, with a panel of 253 commercial European varieties phenotyped with DUS protocol (CREA-DC, Italy), and with multi environmental field trials across Europe (Spain, Italy, Hungary) characterized by agronomic traits monitored in VCU protocol. This panel was genotyped with Illumina Infinium iSelect 90K SNP Chip array (Wang et al., 2014a), performing the SNP call with GenomeStudio Software and augmenting the panel with the Global Durum Panel (Mazzucotelli et al., 2020) up to more than 1000 accessions to increase the source of genetic diversity. The panel was then evaluated for population structure and genetic similarity for each genotype, obtaining separate subpopulations showing the different breeding programs and the high percentage of admixed lines due to modern varieties with common ancestors in the genetic pedigree. In order to augment the resolution of the ancestry analysis, future analysis could be performed on haplotypes. In particular, haplotype blocks and haplotype frequencies could be calculated within the population in order to observe haplotype different frequencies across the genotypes coming from different breeding programs and subpopulation groups. As previously mentioned, the Innovar durum panel (253 varieties) was phenotyped for agronomic traits reported in VCU (more connected with yield and disease resistance) and DUS (morphological and plant developmental traits). Phenotypic data were statistically analysed, and BLUEs were obtained to perform GWAS analysis together with genotypic data. The analysis was performed with different statistical models and results showed major QTLs for different traits, with strong effect on the population and correlation between most associated marker with the phenotypic variation. Reviewing the literature, different confidence intervals for some traits included in VCU and DUS protocols were reported in several publications. For example, known peaks were already detected for grain yield connected with thousand kernel weight or grain protein content (Giunta et al., 2018; Graziani et al., 2014, 2014; Peleg et al., 2009; Roncallo et al., 2017), flowering time (*Ppd*) (Graziani et al., 2014, 2014; Milner et al., 2016), plant growth habit (Mengistu et al., 2016), spike architecture (Blanco et al., 2012) and ear coloration (Patil et al., 2013). For each QTL, a confidence interval and putative candidate genes were identified and some of them were strongly linked with the phenotypic traits. This data could be already useful to build KASP markers associated with phenotypic traits and used to better characterize the agronomic traits reported in the DUS and VCU protocols, giving useful molecular markers for marker assisted selection (MAS) in breeding programs. However, these dataset report only one year of analysis and need to be confirmed with second year of data, performing: 1) separated analysis for year 1 and year 2, expecting to consolidate the QTLs for both years, 2) merge the phenotypic data for both years doing a complete analysis considering also the interaction between genotypes, environments and years. GWAS analysis can be augmented also with results coming from haplotype GWAS, where major peaks for each agronomic trait are expected to be detected as well considering the haplotypes.

The VCU protocol proposes the evaluation of different yield related traits, but also disease resistance is monitored for the variety registration procedure. In Unibo, modern durum wheat resistance was tested for different pathogens, but a detailed case study was presented regarding resistance to SBCMV. Starting from the work published by Maccaferri et al., (2011) and Bruschi et al., (manuscript in preparation), a major QTL, *sbm2*, was detected on chromosome 2B conferring resistance to SBCMV in durum wheat. The *sbm2* interval was explored using the reference durum genome Svevo, converting 11 SNP molecular markers from Illumina 90K in KASP markers used on backcross population Meridiano x MeridianoClaudio BCF_4 to fine map the confidence intervals to

approximately 1 Mb, evaluating different candidate genes inside the interval involved in plant-pathogen interaction and defence response. The genetic interval, compared also with the new Svevo Platinum assembly, was explored also by RNAseq analysis on resistant and susceptible cultivars, detecting different genes overexpressed in resistant cultivars in comparison to the susceptible ones. The strongest candidate genes are represented by protein kinases and NBS-LRR, which are present in different copies inside the interval. These genes are going to be edited using CRISPR-CAS9 technology on Svevo genotype (resistant cultivar), to create knock-out mutant to test in controlled environment for SBCMV infection. A further possibility would be to overexpress the candidate genes in susceptible cultivars, such as Kronos, or doing transient silencing using miRNA in resistant cultivars.

As regards yield related traits, the herein thesis reports the analysis on the *GNI-2A* QTL in the biparental segregant population Relief x Iride F_6 and in the Unibo durum panel. The *GNI-2A* QTL is responsible of increased number of fertile florets per spikelet in Iride, Altar-84 and other CYMMIT related lines. By KASP genotyping analysis, the *GNI-2A* haplotype was detected as a rare allele in modern durum varieties. This means that probably the mutation occurred recently in breeding programs, thus further analysis could be exploited as confirmation. In particular, the Tetraploid Global Collection (TGC, 1500 varieties) available in Unibo (Maccaferri et al., 2019) could be genotyped to check for *GNI-2A* haplotype in wild tetraploid genotypes. Further genotyping analysis can be performed also in Innovar durum panel to detect which variety share (or not) the *GNI-2A* haplotype. The most-likely candidate gene on the *GNI-2A* interval is *hox-2*, Homeobox Zip protein belonging to the transcription factor family. The gene is paralogue of *GNI-1*, orthologue of *vrs* in barley, and based on the expression databases (Borrill et al., 2016), its expression is allocated in immature spike and floret primordia, decreasing than after floret and ovary maturation. This expression pattern is similar to the one of GNI-1, its paralogue gene (Sakuma et al., 2019). To confirm the expression pattern, an RNASeq analysis will be performed on different developmental stages of immature spikes (from floret primordia to complete floret formation) to detect the *hox-2* expression pattern in different varieties, with increased floret fertility and wild type inflorescence.

To conclude, the herein reported thesis shows the application of different phenotyping, genotyping and transcriptomic analysis to characterize agronomic and disease resistance traits in modern durum wheat varieties. The information and results obtained will augment available characterization for each variety, identifying informative molecular markers for breeding purposes and QTLs/candidate genes responsible for the main studied agronomic traits. These data can be proposed to the European variety registration offices to augment the information for varietal characterization included in DUS and VCU protocols.

8. Bibliography

Abdel-Aal, E.S.M., Sosulski, F.W., Hucl, P., 1998. Origins, characteristics, and potentials of ancient wheats. Cereal foods world (USA).

Abu-Zaitoun, S.Y., Chandrasekhar, K., Assili, S., Shtaya, M.J., Jamous, R.M., Mallah, O.B., Nashef, K., Sela, H., Distelfeld, A., Alhajaj, N., Ali-Shtayeh, M.S., Peleg, Z., Ben-David, R., 2018. Unlocking the Genetic Diversity within A Middle-East Panel of Durum Wheat Landraces for Adaptation to Semi-arid Climate. Agronomy 8, 233. https://doi.org/10.3390/agronomy8100233

Achard, F., Butruille, M., Madjarac, S., Nelson, P.T., Duesing, J., Laffont, J.L., Nelson, B., Xiong, J., Mikel, M.A., Smith, J.S.C., 2020. Single nucleotide polymorphisms facilitate distinctness-uniformity-stability testing of soybean cultivars for plant variety protection. Crop Science 60, 2280–2303. https://doi.org/10.1002/csc2.20201

Achtar, S., Moualla, M.Y., Kalhout, A., Röder, M.S., MirAli, N., 2010. Assessment of genetic diversity among Syrian durum (Triticum ssp. durum) and bread wheat (Triticum aestivum L.) using SSR markers. Russ J Genet 46, 1320–1326. https://doi.org/10.1134/S1022795410110074

Akhunov, E.D., Akhunova, A.R., Anderson, O.D., Anderson, J.A., Blake, N., Clegg, M.T., Coleman-Derr, D., Conley, E.J., Crossman, C.C., Deal, K.R., Dubcovsky, J., Gill, B.S., Gu, Y.Q., Hadam, J., Heo, H., Huo, N., Lazo, G.R., Luo, M.-C., Ma, Y.Q., Matthews, D.E., McGuire, P.E., Morrell, P.L., Qualset, C.O., Renfro, J., Tabanao, D., Talbert, L.E., Tian, C., Toleno, D.M., Warburton, M.L., You, F.M., Zhang, W., Dvorak, J., 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11, 702. https://doi.org/10.1186/1471-2164-11-702

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19, 1655–1664. https://doi.org/10.1101/gr.094052.109

Allen, A.M., Winfield, M.O., Burridge, A.J., Downie, R.C., Benbow, H.R., Barker, G.L.A., Wilkinson, P.A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., Griffiths, S., Bentley, A.R., Alda, M., Jack, P., Phillips, A.L., Edwards, K.J., 2017. Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). Plant Biotechnol J 15, 390–401. https://doi.org/10.1111/pbi.12635

Alqudah, A.M., Sallam, A., Stephen Baenziger, P., Börner, A., 2020. GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review. Journal of Advanced Research 22, 119–135. https://doi.org/10.1016/j.jare.2019.10.013

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., Jordan, K.W., Golan, G., Deek, J., Ben-Zvi, B., Ben-Zvi, G., Himmelbach, A., MacLachlan, R.P., Sharpe, A.G., Fritz, A., Ben-David, R., Budak, H., Fahima, T., Korol, A., Faris, J.D., Hernandez, A., Mikel, M.A., Levy, A.A., Steffenson, B., Maccaferri, M., Tuberosa, R., Cattivelli, L., Faccioli, P., Ceriotti, A., Kashkush, K., Pourkheirandish, M., Komatsuda, T., Eilam, T., Sela, H., Sharon, A., Ohad, N., Chamovitz, D.A., Mayer, K.F.X., Stein, N., Ronen, G., Peleg, Z., Pozniak, C.J., Akhunov, E.D., Distelfeld, A., 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357, 93–97. https://doi.org/10.1126/science.aan0032

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [WWW Document], n.d. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 10.31.22).

Babraham Bioinformatics - Trim Galore! [WWW Document], n.d. URL https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed 10.31.22).

Barbosa, M.M., Goulart, L.R., Prestes, A.M., Juliatti, F.C., 2001. Genetic control of resistance to soilborne wheat mosaic virus in Brazilian cultivars of Triticum aestivum L. Thell. Euphytica 122, 417–422. https://doi.org/10.1023/A:1012937116394

Bass, C., Hendley, R., Adams, M.J., Hammond-Kosack, K.E., Kanyuka, K., 2006. The *Sbm1* locus conferring resistance to *Soil-borne cereal mosaic virus* maps to a gene-rich region on 5DL in wheat. Genome 49, 1140–1148. https://doi.org/10.1139/g06-064

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Bayer, M.M., Rapazote-Flores, P., Ganal, M., Hedley, P.E., Macaulay, M., Plieske, J., Ramsay, L., Russell, J., Shaw, P.D., Thomas, W., Waugh, R., 2017. Development and Evaluation of a Barley 50k iSelect SNP Array. Frontiers in Plant Science 8.

Bharadwaj, S., Ginoya, S., Tandon, P., Gohel, T.D., Guirguis, J., Vallabh, H., Jevenn, A., Hanouneh, I., 2016. Malnutrition: laboratory markers vs nutritional assessment. Gastroenterol Rep (Oxf) 4, 272–280. https://doi.org/10.1093/gastro/gow013

Blanco, A., Mangini, G., Giancaspro, A., Giove, S., Colasuonno, P., Simeone, R., Signorile, A., De Vita, P., Mastrangelo, A.M., Cattivelli, L., Gadaleta, A., 2012. Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. Mol Breeding 30, 79–92. https://doi.org/10.1007/s11032-011-9600-z

Boden, S.A., Cavanagh, C., Cullis, B.R., Ramm, K., Greenwood, J., Jean Finnegan, E., Trevaskis, B., Swain, S.M., 2015. Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. Nat Plants 1, 14016. https://doi.org/10.1038/nplants.2014.16

Bolser, D., Staines, D.M., Pritchard, E., Kersey, P., 2016. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. Methods Mol. Biol. 1374, 115–140. https://doi.org/10.1007/978-1-4939-3167-5_6

Bonnett, O.T., 1966. Inflorescences of maize, wheat, rye, barley, and oats : their initiation and development / 721. Bulletin (University of Illinois (Urbana-Champaign campus). Agricultural Experiment Station) ; no. 721.

Borrill, P., Ramirez-Gonzalez, R., Uauy, C., 2016. expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. Plant Physiology 170, 2172–2186. https://doi.org/10.1104/pp.15.01667

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527. https://doi.org/10.1038/nbt.3519

Brozynska, M., Furtado, A., Henry, R.J., 2016. Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnol J 14, 1070–1085. https://doi.org/10.1111/pbi.12454

Budge, G.E., Loram, J., Donovan, G., Boonham, N., 2007. RNA2 of Soil-borne cereal mosaic virus is detectable in plants of winter wheat grown from infected seeds. Eur J Plant Pathol 120, 97–102. https://doi.org/10.1007/s10658-007-9194-9

Bull, H., Casao, M.C., Zwirek, M., Flavell, A.J., Thomas, W.T.B., Guo, W., Zhang, R., Rapazote-Flores, P., Kyriakidis, S., Russell, J., Druka, A., McKim, S.M., Waugh, R., 2017. Barley SIX-ROWED SPIKE3 encodes a putative Jumonji C-type H3K9me2/me3 demethylase that represses lateral spikelet fertility. Nat Commun 8, 936. https://doi.org/10.1038/s41467-017-00940-7

Butruille, D.V., Birru, F.H., Boerboom, M.L., Cargill, E.J., Davis, D.A., Dhungana, P., Dill, G.M., Dong, F., Fonseca, A.E., Gardunia, B.W., Holland, G.J., Hong, N., Linnen, P., Nickson, T.E., Polavarapu, N., Pataky, J.K., Popi, J., Stark, S.B., 2015. Maize Breeding in the United States: Views from Within Monsanto, in: Plant Breeding Reviews: Volume 39. John Wiley & Sons, Ltd, pp. 199–282. https://doi.org/10.1002/9781119107743.ch05

Campbell, R.N., 1996. Fungal transmission of plant viruses. Annu Rev Phytopathol 34, 87–108. https://doi.org/10.1146/annurev.phyto.34.1.87

Ceglar, A., Zampieri, M., Toreti, A., Dentener, F., 2019. Observed Northward Migration of Agro-Climate Zones in Europe Will Further Accelerate Under Climate Change. Earth's Future 7, 1088–1101. https://doi.org/10.1029/2019EF001178

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci 4, 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, W., Wellings, C., Chen, X., Kang, Z., Liu, T., 2014. Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*: *Puccinia striiformis* , yellow rust. Molecular Plant Pathology 15, 433–446. https://doi.org/10.1111/mpp.12116

Cockram, J., Jones, H., Norris, C., O'Sullivan, D.M., 2012. Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (Hordeum vulgare ssp. vulgare L.). Theor Appl Genet 125, 1735–1749. https://doi.org/10.1007/s00122-012-1950-3

Cockram, J., White, J., Zuluaga, D.L., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M.J., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P.E., Stein, N., Schulte, D., Steuernagel, B., Marshall, D.F., Thomas, W.T.B., Ramsay, L., Mackay, I., Balding, D.J., AGOUEB Consortium, Waugh, R., O'Sullivan, D.M., 2010. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci U S A 107, 21611–21616. https://doi.org/10.1073/pnas.1010179107

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., Shaw, P., Bayer, M., Thomas, W., Marshall, D., Hedley, P., Tondelli, A., Pecchioni, N., Francia, E., Korzun, V., Walther, A., Waugh, R., 2012. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 44, 1388–1392. https://doi.org/10.1038/ng.2447

Conxita, R., Ae, F.Á., Ae, V.L., Martos, A.E., Abdelhamid, R., Ae, J.I., Ae, D., Villegas, A.E., Luis, F., Moral, G.D., n.d. Genetic changes in durum wheat yield components and associated traits in Italian and Spanish varieties during the 20th century.

Cooke, R.J., Reeves, J.C., 2003. Plant genetic resources and molecular markers: variety registration in a new era. Plant Genet. Resour. 1, 81–87. https://doi.org/10.1079/PGR200312

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., Coupland, G., 2007. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316, 1030–1033. https://doi.org/10.1126/science.1141752

da Silva, A.F., Sediyama, T., Borém, A., da Silva, F.L., dos Santos Silva, F.C., Bezerra, A.R.G., 2017. Registration and Protection of Cultivars, in: Lopes da Silva, F., Borém, A., Sediyama, T., Ludke, W.H. (Eds.), Soybean Breeding. Springer International Publishing, Cham, pp. 427–440. https://doi.org/10.1007/978-3-319-57433-2_23

De Santis, M.A., Giuliani, M.M., Giuzio, L., De Vita, P., Lovegrove, A., Shewry, P.R., Flagella, Z., 2017. Differences in gluten protein composition between old and modern durum wheat genotypes in relation to 20th century breeding in Italy. European Journal of Agronomy 87, 19–29. https://doi.org/10.1016/j.eja.2017.04.003

Decroës, A., Calusinska, M., Delfosse, P., Bragard, C., Legrève, A., 2019. First Draft Genome Sequence of a Polymyxa Genus Member, Polymyxa betae, the Protist Vector of Rhizomania. Microbiol Resour Announc 8, e01509-18. https://doi.org/10.1128/MRA.01509-18

Dixon, L.E., Farré, A., Finnegan, E.J., Orford, S., Griffiths, S., Boden, S.A., 2018. Developmental responses of bread wheat to changes in ambient temperature following deletion of a locus that includes FLOWERING LOCUS T1. Plant Cell Environ 41, 1715–1725. https://doi.org/10.1111/pce.13130

Dodds, P.N., Rathjen, J.P., 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet 11, 539–548. https://doi.org/10.1038/nrg2812

Doebley, J.F., Gaut, B.S., Smith, B.D., 2006. The molecular genetics of crop domestication. Cell 127, 1309–1321. https://doi.org/10.1016/j.cell.2006.12.006

Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin | Oxford Academic [WWW Document], n.d. URL https://academic.oup.com/book/11300 (accessed 10.8.22).

Dubcovsky, J., Dvorak, J., 2007. Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. Science 316, 1862–1866. https://doi.org/10.1126/science.1143986

Dubey, S.N., Brown, C.M., Hooker, A.L., 1970. Inheritance of Field Reaction to Soil-Borne Wheat Mosaic Virus1. Crop Science 10, cropsci1970.0011183X001000010034x. https://doi.org/10.2135/cropsci1970.0011183X001000010034x

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4, 1184–1191. https://doi.org/10.1038/nprot.2009.97

El Hassouni, K., Belkadi, B., Filali-Maltouf, A., Tidiane-Sall, A., Al-Abdallat, A., Nachit, M., Bassi, F.M., 2019. Loci Controlling Adaptation to Heat Stress Occurring at the Reproductive Stage in Durum Wheat. Agronomy 9, 414. https://doi.org/10.3390/agronomy9080414

Elsen, A.V., Ayerdi Gotor, A., Di Vicente, C., Traon, D., Gennatas, J., Amat, L., Negri, V., Chable, V., 2013. Plant breeding for an EU bio-based economy. (Research Report). auto-saisine.

European Union Intellectual Property Office., Community Plant Variety Office., 2022. Impact of the Community Plant Variety Rights system on the EU economy and the environment. Publications Office, LU.

Evolution of crop plants / [WWW Document], 1995.

Filippo, M., Bassi, F., Nachit, M., 2019. Genetic Gain for Yield and Allelic Diversity over 35 Years of Durum Wheat Breeding at ICARDA. Crop Breeding, Genetics and Genomics 1, e190004. https://doi.org/10.20900/cbgg20190004

Finnegan, E.J., Ford, B., Wallace, X., Pettolino, F., Griffin, P.T., Schmitz, R.J., Zhang, P., Barrero, J.M., Hayden, M.J., Boden, S.A., Cavanagh, C.A., Swain, S.M., Trevaskis, B., 2018. Zebularine treatment is associated with deletion of FT-B1 leading to an increase in spikelet number in bread wheat. Plant, Cell & Environment 41, 1346–1360. https://doi.org/10.1111/pce.13164

Foulkes, M.J., Sylvester-Bradley, R., Worland, A.J., Snape, J.W., 2004. Effects of a photoperiod-response gene Ppd-D1 on yield potential and drought resistance in UK winter wheat. Euphytica 135, 63–73. https://doi.org/10.1023/B:EUPH.0000009542.06773.13

“Framework for the introduction of Plant Breeder’s Rights, Guidance for practical implementation” | Naktuinbouw [WWW Document], n.d. URL https://www.naktuinbouw.com/framework-introduction-plant-breeders-rights-guidance-practical-implementation (accessed 10.9.22).

Gauley, A., Boden, S.A., 2019. Genetic pathways controlling inflorescence architecture and development in wheat and barley. J. Integr. Plant Biol. 61, 296–309. https://doi.org/10.1111/jipb.12732

Gaut, B.S., 2015. Evolution Is an Experiment: Assessing Parallelism in Crop Domestication and Experimental Evolution: (Nei Lecture, SMBE 2014, Puerto Rico). Molecular Biology and Evolution 32, 1661–1671. https://doi.org/10.1093/molbev/msv105

Geldermann, H., 1975. Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. Theoret. Appl. Genetics 46, 319–330. https://doi.org/10.1007/BF00281673

Gilliland, T.J., Gensollen, V., 2010. Review of the Protocols Used for Assessment of DUS and VCU in Europe – Perspectives, in: Huyghe, C. (Ed.), Sustainable Use of Genetic Diversity in Forage and Turf Breeding. Springer Netherlands, Dordrecht, pp. 261–275. https://doi.org/10.1007/978-90-481-8706-5_37

Giunta, F., De Vita, P., Mastrangelo, A.M., Sanna, G., Motzo, R., 2018. Environmental and Genetic Variation for Yield-Related Traits of Durum Wheat as Affected by Development. Frontiers in Plant Science 9.

Glenn, K.C., Alsop, B., Bell, E., Goley, M., Jenkinson, J., Liu, B., Martin, C., Parrott, W., Souder, C., Sparks, O., Urquhart, W., Ward, J.M., Vicini, J.L., 2017. Bringing New Plant Varieties to Market: Plant Breeding and Selection Practices Advance Beneficial Characteristics while Minimizing Unintended Changes. Crop Science 57, 2906–2921. https://doi.org/10.2135/cropsci2017.03.0199

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327, 812–818. https://doi.org/10.1126/science.1185383

González, F.G., Slafer, G.A., Miralles, D.J., 2003. Floret development and spike growth as affected by photoperiod during stem elongation in wheat. Field Crops Research 81, 29–38. https://doi.org/10.1016/S0378-4290(02)00196-X

Graziani, M., Maccaferri, M., Royo, C., Salvatorelli, F., Tuberosa, R., Graziani, M., Maccaferri, M., Royo, C., Salvatorelli, F., Tuberosa, R., 2014. QTL dissection of yield components and morpho-physiological traits in a durum wheat elite population tested in contrasting thermo-pluviometric conditions. Crop Pasture Sci. 65, 80–95. https://doi.org/10.1071/CP13349

Guarda, G., Padovan, S., Delogu, G., 2004. Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian bread-wheat cultivars grown at different nitrogen levels. European Journal of Agronomy 21, 181–192. https://doi.org/10.1016/j.eja.2003.08.001

Hassani-Pak, K., Singh, A., Brandizi, M., Hearnshaw, J., Parsons, J.D., Amberkar, S., Phillips, A.L., Doonan, J.H., Rawlings, C., 2021. KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. Plant Biotechnology Journal 19, 1670–1678. https://doi.org/10.1111/pbi.13583

Hay, R.K.M., 1999. Physiological Control of Growth and Yield in Wheat: Analysis and Synthesis, in: Smith, D.L., Hamel, C. (Eds.), Crop Yield: Physiology and Processes. Springer, Berlin, Heidelberg, pp. 1–38. https://doi.org/10.1007/978-3-642-58554-8_1

Hedden, P., 2003. The genes of the Green Revolution. Trends in Genetics 19, 5–9. https://doi.org/10.1016/S0168-9525(02)00009-4

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., Salamini, F., 1997. Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. Science 278, 1312–1314.

Huang, M., Liu, X., Zhou, Y., Summers, R.M., Zhang, Z., 2019. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8, giy154. https://doi.org/10.1093/gigascience/giy154

Isidro, J., Álvaro, F., Royo, C., Villegas, D., Miralles, D.J., García del Moral, L.F., 2011. Changes in duration of developmental phases of durum wheat caused by breeding in Spain and Italy during the 20th century and its impact on yield. Ann Bot 107, 1355–1366. https://doi.org/10.1093/aob/mcr063

Jamali, S.H., Cockram, J., Hickey, L.T., 2019. Insights into deployment of DNA markers in plant variety protection and registration. Theor Appl Genet 132, 1911–1929. https://doi.org/10.1007/s00122-019-03348-7

Jantasuriyarat, C., Vales, M.I., Watson, C.J.W., Riera-Lizarazu, O., 2004. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (Triticum aestivum L.). Theor Appl Genet 108, 261–273. https://doi.org/10.1007/s00122-003-1432-8

Jördens, R., 2005. Progress of plant variety protection based on the International Convention for the Protection of New Varieties of Plants (UPOV Convention). World Patent Information 27, 232–243. https://doi.org/10.1016/j.wpi.2005.03.004

Kanyuka, K., Lovell, D.J., Mitrofanova, O.P., Hammond-Kosack, K., Adams, M.J., 2004. A controlled environment test for resistance to Soil-borne cereal mosaic virus (SBCMV) and its use to determine the mode of inheritance of resistance in wheat cv. Cadenza and for screening Triticum monococcum genotypes for sources of SBCMV resistance. Plant Pathology 53, 154–160. https://doi.org/10.1111/j.0032-0862.2004.01000.x

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., Weigel, D., 1999. Activation tagging of the floral inducer FT. Science 286, 1962–1965. https://doi.org/10.1126/science.286.5446.1962

Kaur, B., Mavi, G.S., Gill, M.S., Saini, D.K., 2020. Utilization of KASP technology for wheat improvement. CEREAL RESEARCH COMMUNICATIONS 48, 409–421. https://doi.org/10.1007/s42976-020-00057-6

Keeble-Gagnère, G., Isdale, D., Suchecki, R.-ław, Kruger, A., Lomas, K., Carroll, D., Li, S., Whan, A., Hayden, M., Tibbits, J., 2019. Integrating past, present and future wheat research with Pretzel (preprint). Bioinformatics. https://doi.org/10.1101/517953

Kim, S.H., Kim, D.Y., Yacoubi, I., Seo, Y.W., 2021. Development of single-nucleotide polymorphism markers of salinity tolerance for Tunisian durum wheat using RNA sequencing. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science 71, 28–44. https://doi.org/10.1080/09064710.2020.1843701

Kirby, E.J.M., 1974. Ear development in spring wheat. The Journal of Agricultural Science 82, 437–447. https://doi.org/10.1017/S0021859600051339

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., Araki, T., 1999. A pair of related genes with antagonistic roles in mediating flowering signals. Science 286, 1960–1962. https://doi.org/10.1126/science.286.5446.1960

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., Stein, N., Graner, A., Wicker, T., Tagiri, A., Lundqvist, U., Fujimura, T., Matsuoka, M., Matsumoto, T., Yano, M., 2007. Six-

rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci U S A 104, 1424–1429. https://doi.org/10.1073/pnas.0608580104
Konishi, S., Sasakuma, T., Sasanuma, T., 2010. Identification of novel Mlo family members in wheat and their genetic characterization. Genes Genet Syst 85, 167–175. https://doi.org/10.1266/ggs.85.167
Koppolu, R., Anwar, N., Sakuma, S., Tagiri, A., Lundqvist, U., Pourkheirandish, M., Rutten, T., Seiler, C., Himmelbach, A., Ariyadasa, R., Youssef, H.M., Stein, N., Sreenivasulu, N., Komatsuda, T., Schnurbusch, T., 2013. Six-rowed spike4 (Vrs4) controls spikelet determinacy and row-type in barley. Proc Natl Acad Sci U S A 110, 13198–13203. https://doi.org/10.1073/pnas.1221950110
Korte, A., Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9, 29. https://doi.org/10.1186/1746-4811-9-29
Kucek, L.K., Veenstra, L.D., Amnuaycheewa, P., Sorrells, M.E., 2015. A Grounded Guide to Gluten: How Modern Genotypes and Processing Impact Wheat Sensitivity. Comprehensive Reviews in Food Science and Food Safety 14, 285–302. https://doi.org/10.1111/1541-4337.12129
Kühne, T., 2009. Soil-borne viruses affecting cereals—Known for long but still a threat. Virus Research 141, 174–183. https://doi.org/10.1016/j.virusres.2008.05.019
Kumar, J., Pratap, A., Solanki, R.K., Gupta, D.S., Goyal, A., Chaturvedi, S.K., Nadarajan, N., Kumar, S., 2012. Genomic resources for improving food legume crops. The Journal of Agricultural Science 150, 289–318. https://doi.org/10.1017/S0021859611000554
Liller, C.B., Walla, A., Boer, M.P., Hedley, P., Macaulay, M., Effgen, S., von Korff, M., van Esse, G.W., Koornneef, M., 2017. Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. Theor Appl Genet 130, 269–281. https://doi.org/10.1007/s00122-016-2807-y
Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z., 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genetics 12, e1005767. https://doi.org/10.1371/journal.pgen.1005767
Lundqvist, U., Lundqvist, A., 1988. Induced intermedium mutants in barley: origin, morphology and inheritance. Hereditas 108, 13–26. https://doi.org/10.1111/j.1601-5223.1988.tb00677.x
Lynch, J.P., Doyle, D., McAuley, S., McHardy, F., Danneels, Q., Black, L.C., White, E.M., Spink, J., 2017. The impact of variation in grain number and individual grain weight on winter wheat yield in the high yield potential environment of Ireland. European Journal of Agronomy 87, 40–49. https://doi.org/10.1016/j.eja.2017.05.001
Maccaferri, M., Francia, R., Ratti, C., Rubies-Autonell, C., Colalongo, C., Ferrazzano, G., Tuberosa, R., Sanguineti, M.C., 2012. Genetic analysis of Soil-Borne Cereal Mosaic Virus response in durum wheat: evidence for the role of the major quantitative trait locus QSbm.ubo-2BS and of minor quantitative trait loci. Mol Breeding 29, 973–988. https://doi.org/10.1007/s11032-011-9673-8
Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V.M., Milner, S.G., Himmelbach, A., Mascher, M., Bagnaresi, P., Faccioli, P., Cozzi, P., Lauria, M., Lazzari, B., Stella, A., Manconi, A., Gnocchi, M., Moscatelli, M., Avni, R., Deek, J., Biyiklioglu, S., Frascaroli, E., Corneti, S., Salvi, S., Sonnante, G., Desiderio, F., Marè, C., Crosatti, C., Mica, E., Özkan, H., Kilian, B., De Vita, P., Marone, D., Joukhadar, R., Mazzucotelli, E., Nigro, D., Gadaleta, A., Chao, S., Faris, J.D., Melo, A.T.O., Pumphrey, M., Pecchioni, N., Milanesi, L., Wiebe, K., Ens, J., MacLachlan, R.P., Clarke, J.M., Sharpe, A.G., Koh, C.S., Liang, K.Y.H., Taylor, G.J., Knox, R., Budak, H., Mastrangelo, A.M., Xu, S.S., Stein, N., Hale, I., Distelfeld, A., Hayden, M.J., Tuberosa, R., Walkowiak, S., Mayer, K.F.X., Ceriotti, A., Pozniak, C.J., Cattivelli, L., 2019a. Durum wheat genome highlights past domestication signatures and future improvement targets. Nat Genet 51, 885–895. https://doi.org/10.1038/s41588-019-0381-3
Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V.M., Milner, S.G., Himmelbach, A., Mascher, M., Bagnaresi, P., Faccioli, P., Cozzi, P., Lauria, M., Lazzari, B., Stella, A., Manconi, A., Gnocchi, M., Moscatelli, M., Avni, R., Deek, J., Biyiklioglu, S., Frascaroli, E., Corneti, S., Salvi, S., Sonnante, G., Desiderio, F., Marè, C., Crosatti, C., Mica, E., Özkan, H., Kilian, B., De Vita, P., Marone, D., Joukhadar, R., Mazzucotelli, E., Nigro, D., Gadaleta, A., Chao, S., Faris, J.D., Melo, A.T.O., Pumphrey, M., Pecchioni, N., Milanesi, L., Wiebe, K., Ens, J., MacLachlan, R.P., Clarke, J.M., Sharpe, A.G., Koh, C.S., Liang, K.Y.H., Taylor, G.J., Knox, R., Budak, H., Mastrangelo, A.M., Xu, S.S., Stein, N., Hale, I., Distelfeld, A., Hayden, M.J., Tuberosa, R.,

Walkowiak, S., Mayer, K.F.X., Ceriotti, A., Pozniak, C.J., Cattivelli, L., 2019b. Durum wheat genome highlights past domestication signatures and future improvement targets. Nat Genet 51, 885–895. https://doi.org/10.1038/s41588-019-0381-3

Maccaferri, M., Ratti, C., Rubies-Autonell, C., Vallega, V., Demontis, A., Stefanelli, S., Tuberosa, R., Sanguineti, M.C., 2011a. Resistance to Soil-borne cereal mosaic virus in durum wheat is controlled by a major QTL on chromosome arm 2BS and minor loci. Theor Appl Genet 123, 527–544. https://doi.org/10.1007/s00122-011-1605-9

Maccaferri, M., Ratti, C., Rubies-Autonell, C., Vallega, V., Demontis, A., Stefanelli, S., Tuberosa, R., Sanguineti, M.C., 2011b. Resistance to Soil-borne cereal mosaic virus in durum wheat is controlled by a major QTL on chromosome arm 2BS and minor loci. Theor Appl Genet 123, 527–544. https://doi.org/10.1007/s00122-011-1605-9

Maccaferri, M., Ricci, A., Salvi, S., Milner, S.G., Noli, E., Martelli, P.L., Casadio, R., Akhunov, E., Scalabrin, S., Vendramin, V., Ammar, K., Blanco, A., Desiderio, F., Distelfeld, A., Dubcovsky, J., Fahima, T., Faris, J., Korol, A., Massi, A., Mastrangelo, A.M., Morgante, M., Pozniak, C., N'Diaye, A., Xu, S., Tuberosa, R., 2015. A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. Plant Biotechnol. J. 13, 648–663. https://doi.org/10.1111/pbi.12288

Maccaferri, M., Sanguineti, M.C., Corneti, S., Ortega, J.L.A., Salem, M.B., Bort, J., DeAmbrogio, E., del Moral, L.F.G., Demontis, A., El-Ahmed, A., Maalouf, F., Machlab, H., Martos, V., Moragues, M., Motawaj, J., Nachit, M., Nserallah, N., Ouabbou, H., Royo, C., Slama, A., Tuberosa, R., 2008. Quantitative trait loci for grain yield and adaptation of durum wheat (Triticum durum Desf.) across a wide range of water availability. Genetics 178, 489–511. https://doi.org/10.1534/genetics.107.077297

Maccaferri, M., Sanguineti, M.C., Demontis, A., El-Ahmed, A., Garcia del Moral, L., Maalouf, F., Nachit, M., Nserallah, N., Ouabbou, H., Rhouma, S., Royo, C., Villegas, D., Tuberosa, R., 2011c. Association mapping in durum wheat grown across a broad range of water regimes. Journal of Experimental Botany 62, 409–438. https://doi.org/10.1093/jxb/erq287

Marcotuli, I., Gadaleta, A., Mangini, G., Signorile, A.M., Zacheo, S.A., Blanco, A., Simeone, R., Colasuonno, P., 2017. Development of a High-Density SNP-Based Linkage Map and Detection of QTL for β-Glucans, Protein Content, Grain Yield per Spike and Heading Time in Durum Wheat. International Journal of Molecular Sciences 18, 1329. https://doi.org/10.3390/ijms18061329

Maurer, A., Draba, V., Pillen, K., 2016. Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. J Exp Bot 67, 2507–2518. https://doi.org/10.1093/jxb/erw070

Mazzucotelli, E., Sciara, G., Mastrangelo, A.M., Desiderio, F., Xu, S.S., Faris, J., Hayden, M.J., Tricker, P.J., Ozkan, H., Echenique, V., Steffenson, B.J., Knox, R., Niane, A.A., Udupa, S.M., Longin, F.C.H., Marone, D., Petruzzino, G., Corneti, S., Ormanbekova, D., Pozniak, C., Roncallo, P.F., Mather, D., Able, J.A., Amri, A., Braun, H., Ammar, K., Baum, M., Cattivelli, L., Maccaferri, M., Tuberosa, R., Bassi, F.M., 2020. The Global Durum Wheat Panel (GDP): An International Platform to Identify and Exchange Beneficial Alleles. Front. Plant Sci. 11, 569905. https://doi.org/10.3389/fpls.2020.569905

Mengistu, D.K., Kidane, Y.G., Catellani, M., Frascaroli, E., Fadda, C., Pè, M.E., 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 Biotechnology Journal 14, 1800–1812. https://doi.org/10.1111/pbi.12538

Miflin, B., n.d. Crop improvement in the 21st century 8.

Milner, S.G., Maccaferri, M., Huang, B.E., Mantovani, P., Massi, A., Frascaroli, E., Tuberosa, R., Salvi, S., 2016. A multiparental cross population for mapping QTL for agronomic traits in durum wheat (Triticum turgidum ssp. durum). Plant Biotechnology Journal 14, 735–748. https://doi.org/10.1111/pbi.12424

Mitchell-Olds, T., 2010. Complex-trait analysis in plants. Genome Biology 11, 113. https://doi.org/10.1186/gb-2010-11-4-113

Motzo, R., Fois, S., Giunta, F., 2004. Relationship between grain yield and quality of durum wheats from different eras of breeding. Euphytica 140, 147–154. https://doi.org/10.1007/s10681-004-2034-5

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., Buckler, E.S., 2009. Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21, 2194–2202. https://doi.org/10.1105/tpc.109.068437

Nalam, V.J., Vales, M.I., Watson, C.J.W., Kianian, S.F., Riera-Lizarazu, O., 2006. Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (Triticum turgidum L.). Theor Appl Genet 112, 373–381. https://doi.org/10.1007/s00122-005-0140-y

Narasimhamoorthy, B., Gill, B.S., Fritz, A.K., Nelson, J.C., Brown-Guedira, G.L., 2006. Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population. Theor Appl Genet 112, 787–796. https://doi.org/10.1007/s00122-005-0159-0

Neelam, K., Brown-Guedira, G., Huang, L., 2013. Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus Lr21. Mol Breeding 31, 233–237. https://doi.org/10.1007/s11032-012-9773-0

Newell, M.A., Jannink, J.-L., 2014. Genomic selection in plant breeding. Methods Mol Biol 1145, 117–130. https://doi.org/10.1007/978-1-4939-0446-4_10

Newton, A.C., Flavell, A.J., George, T.S., Leat, P., Mullholland, B., Ramsay, L., Revoredo-Giha, C., Russell, J., Steffenson, B.J., Swanston, J.S., Thomas, W.T.B., Waugh, R., White, P.J., Bingham, I.J., 2011. Crops that feed the world 4. Barley: a resilient crop?: Strengths and weaknesses in the context of food security. Food Security 3, 141–178. https://doi.org/10.1007/s12571-011-0126-3

Noleppa, S., Cartsburg, M., 2021. The socio-economic and environmental values of plant breeding in the EU and selected EU member states An ex-post evaluation and ex-ante assessment considering the 327.

Ochagavía, H., Prieto, P., Savin, R., Griffiths, S., Slafer, G., 2018. Dynamics of leaf and spikelet primordia initiation in wheat as affected by Ppd-1a alleles under field conditions. J Exp Bot 69, 2621–2631. https://doi.org/10.1093/jxb/ery104

Options Méditerranéennes en ligne - Collection numérique - Evolution of durum wheat breeding in Italy [WWW Document], n.d. URL https://om.ciheam.org/article.php?IDPDF=00007070 (accessed 10.8.22).

Ordon, F., Habekuss, A., Kastirr, U., Rabenstein, F., Kühne, T., 2009. Virus Resistance in Cereals: Sources of Resistance, Genetics and Breeding. Journal of Phytopathology 157, 535–545. https://doi.org/10.1111/j.1439-0434.2009.01540.x

Patil, R.M., Tamhankar, S.A., Oak, M.D., Raut, A.L., Honrao, B.K., Rao, V.S., Misra, S.C., 2013. Mapping of QTL for agronomic traits and kernel characters in durum wheat (Triticum durum Desf.). Euphytica 190, 117–129. https://doi.org/10.1007/s10681-012-0785-y

Peleg, Z., Cakmak, I., Ozturk, L., Yazici, A., Jun, Y., Budak, H., Korol, A.B., Fahima, T., Saranga, Y., 2009. Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat x wild emmer wheat RIL population. Theor Appl Genet 119, 353–369. https://doi.org/10.1007/s00122-009-1044-z

Pfeiffer, W.H., Sayre, K.D., Reynolds, M.P., 2000. Enhancing genetic grain yield potential and yield stability in durum wheat. Enhancing genetic grain yield potential and yield stability in durum wheat. 83–93.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., Pachter, L., 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14, 687–690. https://doi.org/10.1038/nmeth.4324

Pourkheirandish, M., Dai, F., Sakuma, S., Kanamori, H., Distelfeld, A., Willcox, G., Kawahara, T., Matsumoto, T., Kilian, B., Komatsuda, T., 2018. On the Origin of the Non-brittle Rachis Trait of Domesticated Einkorn Wheat. Frontiers in Plant Science 8.

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., Azhaguvel, P., Sakuma, S., Dhanagond, S., Sharma, R., Mascher, M., Himmelbach, A., Gottwald, S., Nair, S.K., Tagiri, A., Yukuhiro, F., Nagamura, Y., Kanamori, H., Matsumoto, T., Willcox, G., Middleton, C.P., Wicker, T., Walther, A., Waugh, R., Fincher, G.B., Stein, N., Kumlehn, J., Sato, K., Komatsuda, T., 2015. Evolution of the Grain Dispersal System in Barley. Cell 162, 527–539. https://doi.org/10.1016/j.cell.2015.07.002

Prat, N., Guilbert, C., Prah, U., Wachter, E., Steiner, B., Langin, T., Robert, O., Buerstmayr, H., 2017. QTL mapping of Fusarium head blight resistance in three related durum wheat populations. Theor Appl Genet 130, 13–27. https://doi.org/10.1007/s00122-016-2785-0

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–909. https://doi.org/10.1038/ng1847

Principles of Plant Genetics and Breeding, 2nd Edition | Wiley [WWW Document], n.d. . Wiley.com. URL https://www.wiley.com/en-us/Principles+of+Plant+Genetics+and+Breeding%2C+2nd+Edition-p-9780470664759 (accessed 10.9.22).

Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P., 2000. Association Mapping in Structured Populations. The American Journal of Human Genetics 67, 170–181. https://doi.org/10.1086/302959

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S.J., Cory, A.T., Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout, C.J., Chalhoub, B., Mayer, K.F.X., Benhamed, M., Latrasse, D., Bendahmane, A., International Wheat Genome Sequencing Consortium, Wulff, B.B.H., Appels, R., Tiwari, V., Datla, R., Choulet, F., Pozniak, C.J., Provart, N.J., Sharpe, A.G., Paux, E., Spannagl, M., Bräutigam, A., Uauy, C., 2018. The transcriptional landscape of polyploid wheat. Science 361, eaar6089. https://doi.org/10.1126/science.aar6089

Ramirez-Gonzalez, R.H., Uauy, C., Caccamo, M., 2015. PolyMarker: A fast polyploid primer design pipeline. Bioinformatics 31, 2038–2039. https://doi.org/10.1093/bioinformatics/btv069

Ramsay, L., Comadran, J., Druka, A., Marshall, D.F., Thomas, W.T.B., Macaulay, M., MacKenzie, K., Simpson, C., Fuller, J., Bonar, N., Hayes, P.M., Lundqvist, U., Franckowiak, J.D., Close, T.J., Muehlbauer, G.J., Waugh, R., 2011. INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. Nat Genet 43, 169–172. https://doi.org/10.1038/ng.745

Ratti, C., Budge, G., Ward, L., Clover, G., Rubies-Autonell, C., Henry, C., 2004. Detection and relative quantitation of Soil-borne cereal mosaic virus (SBCMV) and Polymyxa graminis in winter wheat using real-time PCR (TaqMan®). Journal of Virological Methods 122, 95–103. https://doi.org/10.1016/j.jviromet.2004.08.013

Ratti, C., Rubies-Autonell, C., Maccaferri, M., Stefanelli, S., Sanguineti, M.C., Vallega, V., 2006. Reaction of 111 cultivars of Triticum durum Desf. from some the world's main genetic pools to soil-borne cereal mosaic virus / Reaktion von 111 Hartweizensorten (Triticum durum Desf.) aus einigen Haupt-Genpools der Welt gegenüber dem Bodenbürtigen Getreidemosaikvirus. Journal of Plant Diseases and Protection 113, 145–149.

Raz Avni, Moran Nave, Tamar Eilam, Hanan Sela, Chingiz Alekperov, Zvi Peleg, Jan Dvorak, Abraham Korol, Assaf Distelfeld, 2014. Ultra-dense genetic map of durum wheat × wild emmer wheat developed using the 90K iSelect SNP genotyping assay. Molecular breeding 34, 1549–1562. https://doi.org/10.1007/s11032-014-0176-2

Reactions of cultivars of common wheat (Triticum aestivum L.) to Soilborne wheat mosaic virus in northern Italy / Reaktionen von Weizensorten auf den Befall durch SBWMV in Norditalien, n.d. 6.

Redman, G., Noleppa, S., 2017. Mycotoxins: The Hidden Danger in Food and Feed. Andersons Research.

Reynolds, M., Trethowan, R.T., van Ginkel, M., Rajaram, S., 2001. Application of Physiology in Wheat Breeding. pp. 2–10.

Roncallo, P.F., Akkiraju, P.C., Cervigni, G.L., 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. https://doi.org/10.1007/s10681-017-2058-2

Rubies-Autonell, C., Vallega, V., Ratti, C., 2003. Reactions of cultivars of common wheat (Triticum aestivum L.) to Soilborne wheat mosaic virus in northern Italy / Reaktionen von Weizensorten auf den Befall durch SBWMV in Norditalien. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz / Journal of Plant Diseases and Protection 110, 332–336.

Russo, M.A., Ficco, D.B.M., Marone, D., De Vita, P., Vallega, V., Rubies-Autonell, C., Ratti, C., Ferragonio, P., Giovanniello, V., Pecchioni, N., Cattivelli, L., Mastrangelo, A.M., 2012. A major QTL for resistance to soil-borne cereal mosaic virus derived from an old Italian durum wheat cultivar. Journal of Plant Interactions 7, 290–300. https://doi.org/10.1080/17429145.2011.640437

Saade, S., Kutlu, B., Draba, V., Förster, K., Schumann, E., Tester, M., Pillen, K., Maurer, A., 2017. A donor-specific QTL, exhibiting allelic variation for leaf sheath hairiness in a nested association mapping population, is located on barley chromosome 4H. PLOS ONE 12, e0189446. https://doi.org/10.1371/journal.pone.0189446

Sakuma, S., Golan, G., Guo, Z., Ogawa, T., Tagiri, A., Sugimoto, K., Bernhardt, N., Brassac, J., Mascher, M., Hensel, G., Ohnishi, S., Jinno, H., Yamashita, Y., Ayalon, I., Peleg, Z., Schnurbusch, T., Komatsuda, T., 2019. Unleashing floret fertility in wheat through the mutation of a homeobox gene. Proc Natl Acad Sci USA 116, 5182–5187. https://doi.org/10.1073/pnas.1815465116

Sakuma, S., Lundqvist, U., Kakei, Y., Thirulogachandar, V., Suzuki, T., Hori, K., Wu, J., Tagiri, A., Rutten, T., Koppolu, R., Shimada, Y., Houston, K., Thomas, W.T.B., Waugh, R., Schnurbusch, T., Komatsuda, T., 2017. Extreme Suppression of Lateral Floret Development by a Single Amino Acid Change in the VRS1 Transcription Factor1[OPEN]. Plant Physiol 175, 1720–1731. https://doi.org/10.1104/pp.17.01149

Sakuma, S., Pourkheirandish, M., Hensel, G., Kumlehn, J., Stein, N., Tagiri, A., Yamaji, N., Ma, J.F., Sassa, H., Koba, T., Komatsuda, T., 2013. Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. New Phytologist 197, 939–948. https://doi.org/10.1111/nph.12068

Sakuma, S., Schnurbusch, T., 2020. Of floral fortune: tinkering with the grain yield potential of cereal crops. New Phytol 225, 1873–1882. https://doi.org/10.1111/nph.16189

Sarao, N.K., Vikal, Y., Singh, K., Joshi, M.A., Sharma, R.C., 2010. SSR marker-based DNA fingerprinting and cultivar identification of rice (Oryza sativa L.) in Punjab state of India. Plant Genetic Resources 8, 42–44. https://doi.org/10.1017/S1479262109990128

Sarkar, P., Stebbins, G.L., 1956. Morphological Evidence Concerning the Origin of the B Genome in Wheat. American Journal of Botany 43, 297–304. https://doi.org/10.2307/2438947

Semagn, K., Babu, R., Hearne, S., Olsen, M., 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breeding 33, 1–14. https://doi.org/10.1007/s11032-013-9917-x

Serrote, C.M.L., Reiniger, L.R.S., Silva, K.B., Rabaiolli, S.M. dos S., Stefanel, C.M., 2020. Determining the Polymorphism Information Content of a molecular marker. Gene 726, 144175. https://doi.org/10.1016/j.gene.2019.144175

Shewry, P.R., 2009. Wheat. Journal of Experimental Botany 60, 1537–1553. https://doi.org/10.1093/jxb/erp058

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., Muricho, G., 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Sec. 5, 291–317. https://doi.org/10.1007/s12571-013-0263-y

Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.-S., Gill, B.S., Faris, J.D., 2006. Molecular characterization of the major wheat domestication gene Q. Genetics 172, 547–555. https://doi.org/10.1534/genetics.105.044727

Slafer, G.A., Rawson, H.M., 1994. Sensitivity of Wheat Phasic Development to Major Environmental Factors: a Re-Examination of Some Assumptions Made by Physiologists and Modellers. Functional Plant Biol. 21, 393–426. https://doi.org/10.1071/pp9940393

Song, J., Bradeen, J.M., Naess, S.K., Raasch, J.A., Wielgus, S.M., Haberlach, G.T., Liu, J., Kuang, H., Austin-Phillips, S., Buell, C.R., Helgeson, J.P., Jiang, J., 2003. Gene RB cloned from Solanum bulbocastanum confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences 100, 9128–9133. https://doi.org/10.1073/pnas.1533501100

Stamp, P., Visser, R.G.F., 2012. The twenty-first century, the century of plant breeding. Euphytica 186, 585–591. https://doi.org/10.1007/s10681-012-0743-8

Steuernagel, B., Witek, K., Krattinger, S.G., Ramirez-Gonzalez, R.H., Schoonbeek, H., Yu, G., Baggs, E., Witek, A.I., Yadav, I., Krasileva, K.V., Jones, J.D.G., Uauy, C., Keller, B., Ridout, C.J., Wulff, B.B.H., 2020. The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. Plant Physiol. 183, 468–482. https://doi.org/10.1104/pp.19.01273

Suprayogi, Y., Pozniak, C.J., Clarke, F.R., Clarke, J.M., Knox, R.E., 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, 437–448. https://doi.org/10.1007/s00122-009-1050-1

Szabo, A.T., Hammer, K., 1996. Notes on the taxonomy of farro: Triticum monococcum, T. dicoccon and T. spelta. Hulled wheats Proceedings of the First International Workshop on Hulled Wheats, held at Castelvecchio Pascoli, Tuscany, Italy, 21-22 July 1995, 2–40.

Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S., Shimamoto, K., 2007. Hd3a protein is a mobile flowering signal in rice. Science 316, 1033–1036. https://doi.org/10.1126/science.1141753

The genetic dissection of quantitative traits in crops | Semagn | Electronic Journal of Biotechnology [WWW Document], n.d. URL http://www.ejbiotechnology.info/index.php/ejbiotechnology/article/view/v13n5-14/1229 (accessed 10.8.22).

The International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C.J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Sharpe, A.G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A.-A., Himmelbach, A., Balfourier, F., Gutierrez-Gonzalez, J., Hayden, M., Koh, C., Muehlbauer, G., Pasam, R.K., Paux, E., Rigault, P., Tibbits, J., Tiwari, V., Spannagl, M., Lang, D., Gundlach, H., Haberer, G., Mayer, K.F.X., Ormanbekova, D., Prade, V., Šimková, H., Wicker, T., Swarbreck, D., Rimbert, H., Felder, M., Guilhot, N., Kaithakottil, G., Keilwagen, J., Leroy, P., Lux, T., Twardziok, S., Venturini, L., Juhász, A., Abrouk, M., Fischer, I., Uauy, C., Borrill, P., Ramirez-Gonzalez, R.H., Arnaud, D., Chalabi, S., Chalhoub, B., Cory, A., Datla, R., Davey, M.W., Jacobs, J., Robinson, S.J., Steuernagel, B., van Ex, F., Wulff, B.B.H., Benhamed, M., Bendahmane, A., Concia, L., Latrasse, D., Bartoš, J., Bellec, A., Berges, H., Doležel, J., Frenkel, Z., Gill, B., Korol, A., Letellier, T., Olsen, O.-A., Singh, K., Valárik, M., van der Vossen, E., Vautrin, S., Weining, S., Fahima, T., Glikson, V., Raats, D., Číhalíková, J., Toegelová, H., Vrána, J., Sourdille, P., Darrier, B., Barabaschi, D., Cattivelli, L., Hernandez, P., Galvez, S., Budak, H., Jones, J.D.G., Witek, K., Yu, G., Small, I., Melonek, J., Zhou, R., Belova, T., Kanyuka, K., King, R., Nilsen, K., Walkowiak, S., Cuthbert, R., Knox, R., Wiebe, K., Xiang, D., Rohde, A., Golds, T., Čížková, J., Akpinar, B.A., Biyiklioglu, S., Gao, L., N'Daiye, A., Kubaláková, M., Šafář, J., Alfama, F., Adam-Blondon, A.-F., Flores, R., Guerche, C., Loaec, M., Quesneville, H., Condie, J., Ens, J., Maclachlan, R., Tan, Y., Alberti, A., Aury, J.-M., Barbe, V., Couloux, A., Cruaud, C., Labadie, K., Mangenot, S., Wincker, P., Kaur, G., Luo, M., Sehgal, S., Chhuneja, P., Gupta, O.P., Jindal, S., Kaur, P., Malik, P., Sharma, P., Yadav, B., Singh, N.K., Khurana, J.P., Chaudhary, C., Khurana, P., Kumar, V., Mahato, A., Mathur, S., Sevanthi, A., Sharma, N., Tomar, R.S., Holušová, K., Plíhal, O., Clark, M.D., Heavens, D., Kettleborough, G., Wright, J., Balcárková, B., Hu, Y., Salina, E., Ravin, N., Skryabin, K., Beletsky, A., Kadnikov, V., Mardanov, A., Nesterov, M., Rakitin, A., Sergeeva, E., Handa, H., Kanamori, H., Katagiri, S., Kobayashi, F., Nasuda, S., Tanaka, T., Wu, J., Cattonaro, F., Jiumeng, M., Kugler, K., Pfeifer, M., Sandve, S., Xun, X., Zhan, B., Batley, J., Bayer, P.E., Edwards, D., Hayashi, S., Tulpová, Z., Visendi, P., Cui, L., Du, X., Feng, K., Nie, X., Tong, W., Wang, L., 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191. https://doi.org/10.1126/science.aar7191

Tian, H.-L., Wang, F.-G., Zhao, J.-R., Yi, H.-M., Wang, L., Wang, R., Yang, Y., Song, W., 2015. Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. Mol Breed 35, 136. https://doi.org/10.1007/s11032-015-0335-0

Trnka, M., Olesen, J.E., Kersebaum, K.C., Skjelvåg, A.O., Eitzinger, J., Seguin, B., Peltonen-Sainio, P., Rötter, R., Iglesias, A., Orlandini, S., Dubrovský, M., Hlavinka, P., Balek, J., Eckersten, H., Cloppet, E., Calanca, P., Gobin, A., Vučetić, V., Nejedlik, P., Kumar, S., Lalic, B., Mestre, A., Rossi, F., Kozyra, J., Alexandrov, V., Semerádová, D., Žalud, Z., 2011. Agroclimatic conditions in Europe under climate change. Global Change Biology 17, 2298–2318. https://doi.org/10.1111/j.1365-2486.2011.02396.x

Turck, F., Fornara, F., Coupland, G., 2008. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol 59, 573–594. https://doi.org/10.1146/annurev.arplant.59.032607.092755

Turner, A., Beales, J., Faure, S., Dunford, R.P., Laurie, D.A., 2005. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310, 1031–1034. https://doi.org/10.1126/science.1117619

Tzarfati, R., Barak, V., Krugman, T., Fahima, T., Abbo, S., Saranga, Y., Korol, A.B., 2014. Novel quantitative trait loci underlying major domestication traits in tetraploid wheat. Mol Breeding 34, 1613–1628. https://doi.org/10.1007/s11032-014-0182-4

Vallega, V., 1985. Reactions of Italian *Triticum durum* Cultivars to Soilborne Wheat Mosaic. Plant Dis. 69, 64. https://doi.org/10.1094/PD-69-64

Varella, A.C., Zhang, H., Weaver, D.K., Cook, J.P., Hofland, M.L., Lamb, P., Chao, S., Martin, J.M., Blake, N.K., Talbert, L.E., 2019. A Novel QTL in Durum Wheat for Resistance to the Wheat Stem Sawfly Associated with Early Expression of Stem Solidness. G3 (Bethesda) 9, 1999–2006. https://doi.org/10.1534/g3.119.400240

Vita, P.D., Matteu, L., Mastrangelo, A.M., Fonzo, N.D., Cattivelli, L., 2007. Effects of breeding activity on durum wheat traits breed in Italy during the 20th century. Italian Journal of Agronomy 2, 451–462. https://doi.org/10.4081/ija.2007.4s.451

Voss-Fels, K.P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., Chen, T.-W., Zetzsche, H., Seddig, S., Majid Baig, M., Ballvora, A., Frisch, M., Ross, E., Hayes, B.J., Hayden, M.J., Ordon, F., Leon, J., Kage, H., Friedt, W., Stützel, H., Snowdon, R.J., 2019. Breeding improves wheat productivity under contrasting agrochemical input levels. Nat Plants 5, 706–714. https://doi.org/10.1038/s41477-019-0445-5

Waddington, S.R., Osmanzai, M., Yoshida, M., Ransom, J.K., 1987. The yield of durum wheats released in Mexico between 1960 and 1984. The Journal of Agricultural Science 108, 469–477. https://doi.org/10.1017/S002185960007951X

Wang, J., Cogan, N.O.I., Forster, J.W., 2016. Prospects for applications of genomic tools in registration testing and seed certification of ryegrass varieties. Plant Breeding 135, 405–412. https://doi.org/10.1111/pbr.12388

Wang, J., Luo, M.-C., Chen, Z., You, F.M., Wei, Y., Zheng, Y., Dvorak, J., 2013. Aegilops tauschii single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytol 198, 925–937. https://doi.org/10.1111/nph.12164

Wang, J., Zhang, Z., 2021. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Genomics, Proteomics & Bioinformatics, Bioinformatics Commons 19, 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., Mastrangelo, A.M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing Consortium, Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A.R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K.J., Hayden, M., Akhunov, E., 2014a. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12, 787–796. https://doi.org/10.1111/pbi.12183

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., Mastrangelo, A.M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing Consortium, Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A.R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K.J., Hayden, M., Akhunov, E., 2014b. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12, 787–796. https://doi.org/10.1111/pbi.12183

Wang, Y., Yu, H., Tian, C., Sajjad, M., Gao, C., Tong, Y., Wang, X., Jiao, Y., 2017. Transcriptome Association Identifies Regulators of Wheat Spike Architecture. Plant Physiol. 175, 746–757. https://doi.org/10.1104/pp.17.00694

Woodward, J., 2014. Bi-Allelic SNP Genotyping Using the TaqMan® Assay, in: Fleury, D., Whitford, R. (Eds.), Crop Breeding: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 67–74. https://doi.org/10.1007/978-1-4939-0446-4_6

Wu, T.D., Watanabe, C.K., 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Würschum, T., Leiser, W.L., Langer, S.M., Tucker, M.R., Longin, C.F.H., 2018. Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. Theor Appl Genet 131, 2071–2084. https://doi.org/10.1007/s00122-018-3133-3

Xynias, I.N., Mylonas, I., Korpetis, E.G., Ninou, E., Tsaballa, A., Avdikos, I.D., Mavromatis, A.G., 2020. Durum Wheat Breeding in the Mediterranean Region: Current Status and Future Prospects. Agronomy 10, 432. https://doi.org/10.3390/agronomy10030432

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S., Dubcovsky, J., 2006. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proceedings of the National Academy of Sciences 103, 19581–19586. https://doi.org/10.1073/pnas.0607142103

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., Dubcovsky, J., 2003. Positional cloning of the wheat vernalization gene VRN1. Proceedings of the National Academy of Sciences 100, 6263–6268. https://doi.org/10.1073/pnas.0937399100

Yang, C.J., Russell, J., Ramsay, L., Thomas, W., Powell, W., Mackay, I., 2021. Overcoming barriers to the registration of new plant varieties under the DUS system. Commun Biol 4, 302. https://doi.org/10.1038/s42003-021-01840-9

Yang, J., Liu, P., Zhong, K., Ge, T., Chen, L., Hu, H., Zhang, T., Zhang, H., Guo, J., Sun, B., Chen, J., 2022. Advances in understanding the soil-borne viruses of wheat: from the laboratory bench to strategies for disease control in the field. Phytopathol Res 4, 27. https://doi.org/10.1186/s42483-022-00132-2

Youssef, H.M., Eggert, K., Koppolu, R., Alqudah, A.M., Poursarebani, N., Fazeli, A., Sakuma, S., Tagiri, A., Rutten, T., Govind, G., Lundqvist, U., Graner, A., Komatsuda, T., Sreenivasulu, N., Schnurbusch, T., 2017. VRS2 regulates hormone-mediated inflorescence patterning in barley. Nat Genet 49, 157–161. https://doi.org/10.1038/ng.3717

Yu, J.-K., Chung, Y.-S., 2021. Plant Variety Protection: Current Practices and Insights. Genes 12, 1127. https://doi.org/10.3390/genes12081127

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Research 14, 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J.C., Deal, K.R., De Oliveira, R., Choulet, F., Keeble-Gagnère, G., Tibbits, J., Rogers, J., Eversole, K., Appels, R., Gu, Y.Q., Mascher, M., Dvorak, J., Luo, M.-C., 2021. Optical maps refine the bread wheat Triticum aestivum cv. Chinese Spring genome assembly. The Plant Journal 107, 303–314. https://doi.org/10.1111/tpj.15289

Zhu, T., Wang, L., Rodriguez, J.C., Deal, K.R., Avni, R., Distelfeld, A., McGuire, P.E., Dvorak, J., Luo, M.-C., 2019. Improved Genome Sequence of Wild Emmer Wheat Zavitan with the Aid of Optical Maps. G3 Genes|Genomes|Genetics 9, 619–624. https://doi.org/10.1534/g3.118.200902