Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze e Tecnologie Agrarie, Ambientali e Alimentari

Ciclo XXVI

Settore Concorsuale di afferenza: 07/E1

Settore Scientifico disciplinare:AGR/07

A multiparental cross population for mapping QTL for relevant agronomic traits in durum wheat (*Triticum durum* Desf.)

Presentata da: Dott.ssa Sara Giulia Milner

Coordinatore Dottorato :

Chiar.mo Prof. Giovanni Dinelli

Relatore:

Dott. Silvio Salvi

Correlatori:

Chiar.mo Prof. Roberto Tuberosa

Dott. Marco Maccaferri

Esame finale anno 2014

INTRODUCTION

1. Durum wheat

1.1 Botany of the genus *Triticum*

Wheat genus (*Triticum* ssp.) belongs to the Graminaceae family (i.e. Poaceae), *Triticeae* tribe and *Triticinae* subtribe. It includes, among a rich variety of other species, two major crops important for human diet: bread wheat (*Triticum aestivum*) and durum wheat (*Triticum turgidum*, i.e. *Triticum durum* Desf.). Although very similar in many aspects, the two species are well differentiated at genetic (ploidy) and agronomical (main growing areas) levels. Bread and durum wheat are allotetraploid and alloheaxaploid species respectively, with large and complex genomes (Ganal and Röder, 2007).

From a genetic point of view, wheat species represent an allopolyploid series (x = 7) based on genomes A, B, D and G. The series includes:

1) Diploid wheat species (2n = 2x = 14 chromosomes) with genomic formula AA;

2) Allotetraploid species (2n = 4x = 28 chromosomes) with genomic formula AABB (*Triticum dicoccoides*, *Triticum dicoccum* and *Triticum turgidum* ssp. *durum* Desf.) and AAGG (Triticum *timopheevi*);

3) Allohexaploid species (2n=6x=42 chromosomes) and genomic formula AABBDD (*Triticum aestivum* ssp. *spelta* and *Triticum aestivum* ssp. *aestivum*).

The phylogenetic relationships among wild and cultivated wheat species have been extensively studied (Kihara, 1919, 1924). *Triticum* species are usually largely interfertile and the evolutionary history of cultivated wheats witnessed several interspecific hybridization events (Blanco *et al*., 1990; Monneveux *et al*., 2000). The evolutionary history of cultivated wheat species have started around 10,000 – 8,500 years ago with a series of interspecific hybridizations (see Fig. 1) followed by spontaneous chromosome doublings which produced fertile polyploid progeny characterized by bivalent chromosome pairing at meiosis (Ozkan *et al.,* 2001, Shaked *et al.,* 2001).

The discovery of the diploid species that conferred genomes to cultivated wheat species is an important task both for elucidating the actual phylogenetic relationships and for identifying potential sources of beneficial alleles (Blanco *et al*., 1990; Monneveux *et al*., 2000). The A genome is considered as the pivot genome common to all wheat species and derives from an ancestor of the wild wheat *Triticum urartu* (AA genome, 2n = 14. Dvorak *et al,.* 1992). The origin of the B genome is more complex to be traced (Zohary and Feldman, 1962; Talbert *et al.*, 1995).

It has been suggested that the B genome originates from the SS genome of an *Aegilops* species belonging to the *Sitopsis* section and similar (van Slageren, 1994) to the present *Aegelops speltoides* (Sarkar and Stebbins, 1956). The D genome of bread wheat and of *Triticum spelta* was the result of a second independent hybridization (for a review, see Dubcovsky and Dvorak, 2007) between an ancestor of the diploid *Aegilops tauschii* var. *strangulate* (DD genome, McFadden and Sears, 1946) and an allotetraploid wheat. Based on recent investigations (Akhunov *et al.*, 2010) the tetraploid genome of cultivated durum wheat appears homolog to the wild allotetraploid *T. dicoccoides* and to the ancestral domesticated species *T. dicoccum* (emmer wheat) while bread wheat genome shows analogies with *T. aestivum* ssp. *spelta* , i.e. spelt (Slageren, 1994).

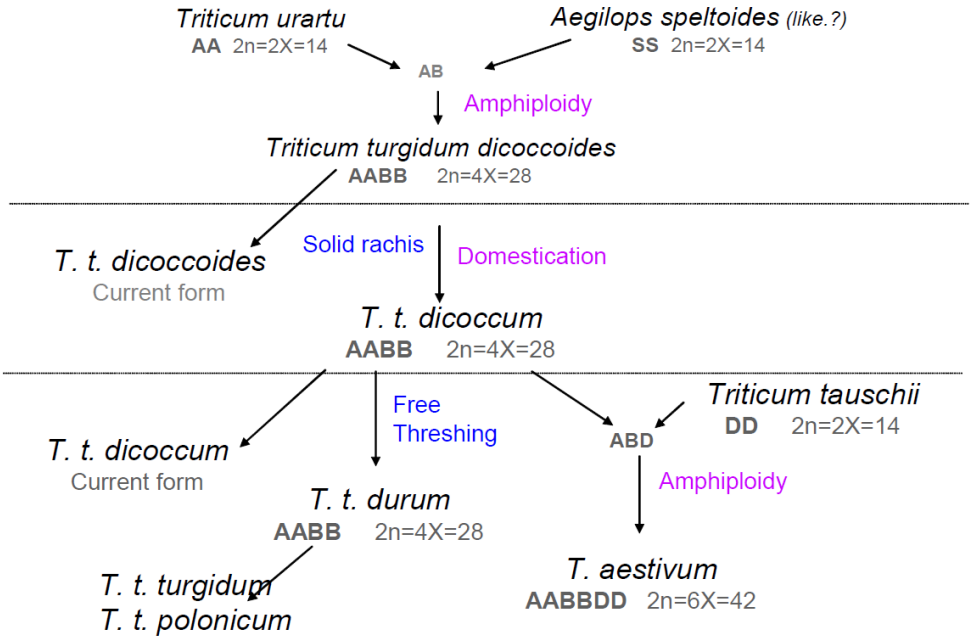


Fig.1. Phylogeny of wheat species.

1.2 Economic importance of durum wheat

Wheat is grown worldwide on more than 216 million hectares, with a world production of 670 million tons in 2012 (http://faostat3.fao.org/), and world's top producers being China, India, Russia and the United States. Therefore, wheat is the third most-produced cereal after maize (well over 800 million tons) and rice (over 700 million tons), occupying more arable land than any other crop.

Durum wheat accounts only for 10% of total wheat-world production but is considered a primary staple crop particularly in the Mediterranean basin, where it is grown across several macro-environments rather differentiated with regards to the thermo-pluviometrical conditions during the crop cycle (Loss and Siddique, 1994). Durum wheat is generally grown under rainfed conditions and is adapted to relatively harsh environments prone to drought stresses (Bozzini, 1988, Araus *et al*., 2002; Condon *et al.,* 2004).

Durum wheat is used for a variety of food products. Its vitreous kernel is composed by 60-80% of carbohydrates (principally starches); 8-16% of proteins; 1.5-2% of fat, 1.5-2% of minerals, 2.2% of fibres, vitamin B and E (Pena, 2002). The majority of durum grown in the Mediterranean basin is milled to form semolina, which is then processed in a variety of food products (pasta, cus-cus, burgur, as well as flat breads).

1.3 Wheat phenology

Although crop development is a continuum process, it is often schematically regarded as a sequence of phenological steps characterized by functional and morphological modifications (Landsberg, 1977). Efforts in modelling wheat development generally focus in defining the phenology related to apex development in relation to the stages most important for particular grain yield components (see Fig. 2). The scheme possess an arbitrary time-scale as the actual scale is in fact affected by the interaction between genetic and environmental factors. Wheat cycle is usually divided into the following growth stages: germination, seedling growth; tillering; stem elongation; booting; ear emergence; flowering; milk development; dough development; ripening. Such stages can be grouped in three main three sub-phases (Slafer and Rawson, 1994): 1) vegetative stage (from seed imbibition to floral initiation 2) early reproductive (in which reproductive organs' differentiation occurs) and 3) late reproductive/maturity stage (from terminal spikelet initiation to end of grain filling). The most wide-spread scale which discretely describes wheat development have been proposed by Zadoks *at al*., 1974. Zadok's growth scale (0-99) is a decimal code comprising two digits. The first indicates one of the ten principal wheat growth stages (0-9 from germination to ripening, respectively) while the second digit subdivides each principal growth stage according to secondary stages of development.

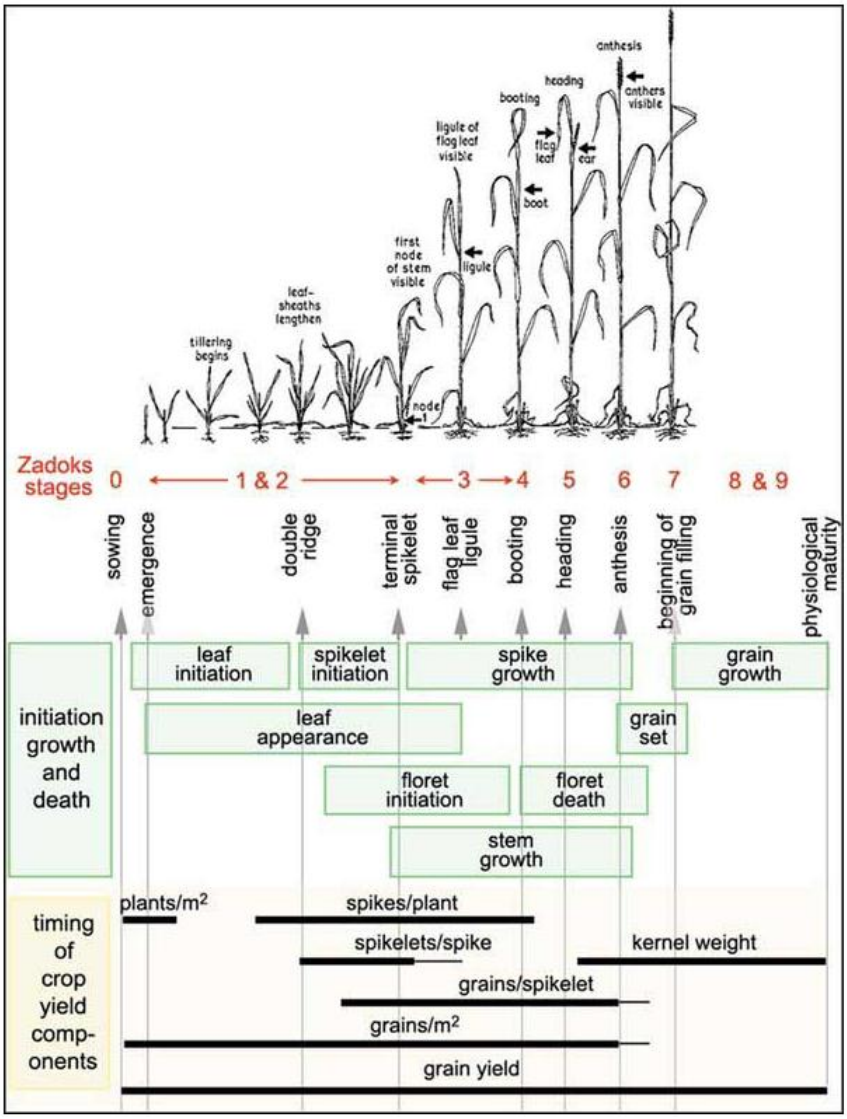


Fig. 2 Phenological stages in wheat development and their relationship to particular grain yield components. Figure From: FAO Irrigated Wheat. Howard M Rawson and Helena Gómez Macpherson.

1.4 Vernalization and photoperiod responses

In wheat species, the control of the irreversible transition from vegetative to reproductive phase (the initiation of the first floral primordia) is crucial event for its adaptation to a wide range of climatic environments and has been heavily targeted during domestication and breeding. The time required for a wheat plant to reach flowering (i.e. precocity or earliness) is the result of several interactions involving both genetic and environmental factors (Brooking, 1996; Robertson *et al*., 1996) and it is generally divided into three components: vernalization (a plant's requirement for a prolonged exposure to cold temperatures in order to initiate/accellerate flowering (Chouard, 1960), photoperiod sensitivity (when flowering time is defined by long

days' sensing) and earliness *per se*, i.e. variability in precocity independently from the mechanisms of vernalization and photoperiod (Kato and Yamagata, 1988; Dubcovsky *et al*., 2007). With regards to the strength of vernalization response, it is possible to define two major classes of grow habit in wheat (Flood and Halloran, 1986): 1) spring wheat (very mild or no response to vernalization, i.e. capable of flowering without vernalization treatment) and 2) winter wheat (i.e. strong response to vernalization). Then, wheat genotypes can be subdivided based on their sensitivity (or no-sensitivity) to photoperiod. Most of cultivated varieties are defined as quantitative long-day plants which flower faster as the day-length increases but do not need a particular day-length to initiate the process (Evans *et al*., 1975) 'Photoperiod insensitive' mutations have then conferred to certain wheat genotypes the early flowering trait, which is advantageous in drought-prone environments as it could allow the plant to complete the life cycles before the likely occurrence of environmental stresses (Law and Worland, 1997; Kato and Yokoyama, 1992). Wheat genetic variability for flowering time largely explains the successful of its worldwide cultivation and heavy influences wheat yield potential in a range of temperate to sub-tropical regions. Investigations on the genetic bases quantitatively determining the length of the vegetative cycle and flowering initiation in are of crucial importance for further improving productivity in wheat. The major genes (as well as their molecular interactions) controlling vernalization and photoperiod response in wheat have been detected (see Trevaskis *et al*., 2007; Distelfeld *et al*., 2009; Kumar *et al*., 2012a). Three main genes associated to vernalization response in wheat have been identified: *VRN1* and *VRN2* on group 5 chromosome and *VRN3* on group 7 chromosome (Yan *et al*. 2003, 2004) and their molecular interactions have been extensively investigated (Tranquilli and Dubcovsky, 2000). *Vrn*-B3 locus on 7BS (Law and Worland, 1997) has been associated to growth habit in wheat and has been identified as an *FT*-like gene (Yan *et al*., 2006). In bread wheat, some major genes for photoperiod response have been characterized and mapped on group 2 chromosome: Ppd-D1, Ppd-B1, and Ppd-A1 (Kato and Yokoyama 1992; Dubcovsky *et al.,* 2006) identified as a *PRR*-like genes (Yan *et al.,* 2006; Bonnin *et al.,* 2008;). In tetraploid wheat, a candidate *Ppd* locus has been detected on chromosome chr 2A (Wilhelm *et al.,* 2009, Maccaferri *et .al.,.* 2008). Loci controlling earliness-per se contribute in fine tuning flowering time and are thought to be internally self-regulated by the plant, independently from variation in light (photoperiod or day length) and vernalization-sensitiveness (Kato and Wada, 1999; Hanocq *et al.,* 2007; Le Gouis *et al*., 2012).

1.5 Yield potential in wheat

Yield is an extremely complex trait which derives from the final balance among different factors determined at different stages of development and it is highly influenced by the environment

(Reynolds *et al.,* 2001). The physiological and numerical yield components which contribute to the overall crop yield could be exemplified by two well-known equations (Hay and Walker, 1989; Slafer, 2003) namely:

1) 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.

From a simpler numerical point of view, the final yield (per m^2) is considered as

2) GY=NG m^{-2} * IGWt

Where NG m^{-2} and IGWt stand for: number of grains per unit area and individual grain weight, respectively.

The 'number of grains per unit of area' is itself determined by many individual sub-components which have been schematically identified as: Spikes m^{-2} (determined by both spikes*$plant^{-1}$ and plants*m^{-2}) and Grains $spike^{-1}$ (determined by grains*$spikelet^{-1}$ and spikelets*$spike^{-1}$) and compensate each other along all the plant development (Slafer and Rawson, 1994). The 'average individual grain weight' has a range of variation inferior to those possessed by NG m^{-2} and a lower correlation to net yield (Slafer and Andrade, 1993).

Crop development is usually described with regards to the phases which seems more critical for determining yield potential. A crucial period for final yield determination has been identified between terminal spikelet initiation and anthesis, around 20-30 days before anthesis (Fisher, 1984; Slafer and Rawson, 1994). At this stage of wheat development, the elongating stem concurs in fact with spike development for the partition of assimilates influencing Grains $spike^{-1}$ (Frederick and Bauer, 1999). As underlined by Fischer (1984), a strategy aimed at increasing yield potential acting exclusively on single numerical yield determinants is often turning to be ineffective, because negative relationship among components exist. Since the magnitude of the parameters describing such relationships is variable, it is in fact impossible to predict net yield. Yield improvement have been often performed seeking for mapping 'yield *per se*' quantitative loci which, as expected, it is considered as a very hard task (Loss and Siddique, 1994).

1.6 Genetic improvement of wheat

World's trend in wheat production has seen a substantial increase during the second half of the 20th century (from 303 million tons in 1966 to 606 million tons in 2006, Evenson and Gollin, 2003) as a consequence of two causes of very different nature: an increase in the overall harvested area which occurred mainly in the first half of the century, followed by the raising importance (from 1955 onwards) of increased grain yield which has been ascribed both to genetic gain in yield potential and to improved agronomic practices (Slafer *et al*., 1994). A considerable boost in genetic gains in wheat yield potential have occurred during the so-called 'Green Revolution' with the release of new improved varieties. The main drivers of such revolution were represented by efforts towards the optimization of crop development and phenology.

Wheat phenological development has been manipulated to reduce competition among tissues in the pre-anthesis allowing a better allocation of assimilates toward spikes following two main strategies.

1) Pursuing a better reallocation of the biomass towards the reproductive organs (higher harvest index)

Such change in the dry matter partitioning have been achieved reducing stem elongation via the introgression in bread wheat of major genes controlling plant height and conferring a semi-dwarf phenotype (*Rht* genes, Richards, 1992a, 1992b). As floral abortion usually coincides with the period of maximum rate of both stem and spike growing (Siddique *et al*., 1989; Youssefian *et al*., 1992) it has been hypothesized that the increased survival of florets (thereby resulting in more whilst smaller grains per spike) observed with dwarf/semi-dwarf cultivars could derive from a reduced competition between the growing spike and the elongating stem (Calderini, Deccer and Slafer, 1995; Isidro *et al*., 2011; Rebetzke *et al*., 2012). The presence of such physiological mechanism has been hypothesized also in the case of durum wheat by Miralles et al. (2002). These changes, in turn, enabled to increase the crop nitrogen fertilization without the negative impact on plant architecture (unbalanced harvest index and stem lodging) (Fischer and Stockman 1986)

2) Delaying anthesis and increasing the photosynthetic activity in pivotal phases

This objective has been principally achieved manipulating crop response to photoperiod and temperature with the deployment of major genes controlling photoperiod-sensitivity (*Ppd*

genes, Turner *et al.* 2005; Foulkes *et al*., 2004) and *Vrn*, vernalisation-response genes (Yan *et al*., 2003, 2006).

As a matter of fact, while breeding programs during the 20th century have contributed to increase the number of grains per unit land area, little or no changes (or even a deployment) in individual grain weight have been observed (Austin *et al*., 1980; Slafer and Andrade1989; Sadras and Lawson, 2011). The negative correlation between number of grains per unit land area and individual grain weight generally found in wheat varieties (Slafer *et al.*, 1996) is thought to be not assignable to feedback regulation as these two components are determined during different developmental stages with only minor temporal overlapping. Understanding the physiological basis of such negative compensation is of pivotal importance in wheat breeding, as this mutual mechanism it is hampering further genetic improvements in terms of grain yield (Slafer *et al.*, 1996). Two possible causes have been summarized by Slafer *et al*. (1996): 1) Competitivity: increasing the grain number per ear reduces the growing ratio of grains themselves as it reduces the available assimilates; 2) intrinsic minor weight potential of additional grains for instance positioned at the top or at the bottom of ear and spikelets. Although the first one is the most widely accepted explanation, Slafer et al. (1996) pointed out that only a finer and deeper physiological investigation about the complex mechanisms determining yield (e.g. crop ability in using radiation, absorbing nutrients and allocating them into competing organs) could lead to further increases in yield potential.

Growing evidence is being accumulated regarding the hypothesis that sink strength is still a critical yield limiting factor in wheat (Miralles et al, 2000; Miralles and Slafe, 2007) and that elite materials are still currently facing unnecessary floret abortion due to non-optimal exploitation of their photosynthetic capacity (Borras *et al, 2004*). Therefore, improving the source-sink balance could represent a promising strategy for raising RUE, biomass and final grain yield (Calderini *et al*., 1997; Reynolds *et al*., 2001, 2005; Shearman *et al*., 2005, Richards, 1996; Slafer *et al*., 1996). As suggested by Reynolds *et al.* (2007), sink-source (SS) related traits could be grouped in three different classes: 1) Finer-tuning of crop phenological patterns (e.g. focussing on early-*per se* QTL. 2) Assimilation capacity of the crop until shortly after anthesis. 3) Partitioning of assimilates to competing sinks during spike growth.

Focussing on the latest class, a larger genetic plasticity both in terms of kernel size and number of kernel per ear could likely contribute to efficiently accommodate an increased carbon remobilization from source tissues. Such partitioning trade-offs should be accompanied by a

parallel improvement in both stem strength and root anchorage system in order to prevent lodging and thus yield losses (Reynolds *et al.,* 2009).

Grain protein quantity in wheat, besides determining grain nutritional value, it is a trait affecting the end-use quality of bread and durum wheat and is controlled by a complex genetic system. Modern high-yielding wheat cultivars are characterized by low grain protein content (GPC), due to its generally negative correlation with grain yield components (Zanetti *et al*., 2001; Blanco *et al*., 2002, 2006).Grain weight and yield penalties have been identified also as associated to the first cloned wheat gene responsible for variation in GPC, *GPC-B1* (Uauy *et al*., 2006).

2. The dissection of quantitative traits

Most traits of interest in agricultural genetics show a continuous range of variation and a complex mode of inheritance which appear to not follow Mendelian inheritance patterns. As suggested by the multiple-factor hypothesis by Nillson-Ehle (1909) and East (1916), they are thought to be controlled by the cumulative action of several genes, along with environmental factors and epistatic interactions (Tanksley, 1993; Falconer and Mackay, 1996; Lynch and Walsh, 1998).

A quantitative trait locus (QTL, Geldermann, 1975) is a statistically identified genomic region hypothetically associated with genetic variation of a complex trait. The dissection of a quantitative trait implies the identification of the QTL contributing to the expression of a quantitative phenotype, by means of molecular markers statistically associated with the variation at the phenotypic level (Lynch & Walsh, 1998). Genotype-phenotype associations are detected when different classes of marker alleles show statistically different trait values, as a consequence of linkage disequilibrium between tested marker positions and genomic loci that underpin trait variation (Lynch and Walsh, 1998).

The basic approach underlying QTL analysis is the comparison of a model assuming no QTL (H_0: no QTL at the test position) with the alternative hypothesis expecting a QTL to be present in genetic linkage to the target marker/s. A number of statistical methods intended to test the validity of H_0 have been proposed and they are generally subdivided into two main categories (for a review, Broman, 2001; Wurshum, 2012). The first one includes those approaches requiring the development of *ad hoc* experimental populations (this category is usually named QTL analysis by linkage mapping), the second comprises methods relying on existing

populations (natural populations, germplasm collection, breeding materials, etc.) and exploits their linkage disequilibrium (this category is usually referred to as linkage disequilibrium mapping or association mapping).

The results of QTL analysis contribute information on the genetic architecture of a target complex trait, including the number and position of quantitative loci detected as responsible for trait variation, the relative portion of genetic variance explained, the magnitude and direction of effect of each locus (i.e. the parental source of beneficial alleles at each QTL). More advanced approaches allow as well to identify QTL interactions (epistasis) and/or genotype x environment interactions (Bradshaw, 1996).

The success of a QTL mapping experiment depends on many factors of different nature (e.g. the type and size of chosen mapping population, the complexity of the investigated trait, the quality of the linkage map). As the genetic variation of a quantitative trait could involve a very large number of loci with minor effect (Kearsey and Farquhar, 1998), a high trait hereditability is crucial for reliable QTL results (Bradbury *et al.*, 2011). To limit the weight of additional sources of variation on the target trait, the design of a QTL experiment generally requires to collect phenotypic data from replicated field experiments with large sample of individuals and to allow the observation of an adequate number of recombinants.

In last decades, the improvement in high-throughput genetic marker systems and the development of high-density linkage maps and specific analytical software, have resulted in the successful application of QTL analysis for deciphering complex traits in the field of crop genetics (Doerge *et al.,* 1997; Mackay, 2009).

The mapping of agriculture-relevant QTL could also lead to the applicative goal of identifying markers to be used as indirect selection tools in marker-assisted breeding, (Semagn *et al*., 2006; Bernardo, 2008, see Chapter 3).

2.1 Linkage-based QTL mapping

QTL mapping with experimental crosses can be further subdivided in single marker analysis (SMA, Sax, 1923) and interval mapping (IM, Lander and Botstein, 1989). More advanced methods enable to contemporarily consider multiple QTL comprise composite interval mapping (CIM, Zeng, 1993; Jansen and Stam, 1993), multiple interval mapping (Kao *et al.*, 1999).

Once obtained the required two sets of data from the same segregating population (phenotypic measurements of a quantitative trait and a molecular marker dataset), the simplest method for

assessing the difference in the phenotypic means with respect to a single-marker genotype is represented by *F*-statistics ANOVA test (or *t*-test, if only two possible genotypes are determined by the population type, Sax, 1923; Soller *et al*., 1976). The sampled population is therefore split into groups according to the existing genotype classes at each marker position and a null hypothesis (H_0 = the mean of the trait value is independent of the genotype at a particular marker) is tested. If previous knowledge about other marker loci associated to the trait of interest is present, the statistical model could include such constant set of background markers and test the datasets for further associations by multiple regression (Manly and Olson, 1999).
Important constraints of SMA have been underlined by many authors (Lander and Botstein, 1989; Manly and Olson, 1999; Broman, 2001). First of all, although being a simple method for pointing the existence of potential QTL, SMA hardly provides a reliable estimation of QTL location since the QTL effect and QTL location are unable to be estimated separately. SMA is quite sensitive both to unbalanced datasets (individuals with missing genotypes need often to be discarded unless a mixed model is utilized) and to uneven distribution of markers which can cause a strong decrease in QTL detection power.

More sophisticated methods for mapping QTL with improved location accuracy, rely both on the simultaneous use of multiple marker information and on the availability of a genetic map estimated in the population under investigation. Lander and Botstein (1989) firstly proposed a simple interval mapping approach, SIM, where the existence of a putative QTL is tested conditional upon flanking marker genotypes at incremental map intervals defined by ordered pairs of markers. Lander and Botstein's IM is a likelihood-based method where a likelihood-of-odd score (LOD) is estimated by an EM algorithm at each increment across the genetic map to evaluate the likelihood of the null hypothesis compared to a single-QTL model. Where the LOD profile exceeds a genome-wide significance threshold, a QTL is declared at the interval with the highest LOD score. IM is considered as more efficient than SMA in locating quantitative loci, as QTL effects are estimated after incrementally fixing the position of an hypothetical QTL. Haley and Knott (1992) and Martinez and Curnow (1992) both proposed a reliable approximation to the described computationally-demanding interval mapping based on maximum likelihood. Phenotypic values were then regressed onto expected genetic coefficients of a putative QTL at tested positions, by means of a least square equivalence. The approximation has been proved as very close to IM results based on likelihood profile by Haley and Knott (1992) and Rebai *et al.* (1995). However, Xu (1995) underlined that the model residual variance could be overestimated and presented a correction in order to not affect the IM QTL detection power.

Although the additional information supplied by the genetic map enables SIM to be considered as a more powerful method than ANOVA for mapping QTL (Lander and Botstein, 1989; Haley and Knott, 1992; Zeng, 1994), SIM is still a single-QTL model and therefore presents its own limitations. With SIM the effects of other possible QTL on the same linkage group are ignored, thus providing a biased estimation of QTL effect and position when such multiple quantitative loci exist (Haley and Knott, 1992; Knott and Haley, 1992; Martinez and Curnow, 1992; Zeng, 1994). Moreover, the likelihood profile in a certain interval may exceed the significance threshold even when there is no QTL within the target markers (a phenomenon known as 'ghost peak'), due to the effect of other QTL linked nearby to the interval of interest. Motivated by those issues, Jansen and Stam (1993) and Zeng (1993) proposed to introduce as covariates additional flanking markers into the likelihood function. Such model QTL mapping model resulted thus adjusted for the confounding effects of loci located outside a currently scanned interval. Both Jansen's multiple QTL mapping and Zeng's composite interval mapping (CIM) evaluate the existence of a QTL at multiple analysis points across a given interval, while simultaneously fitting partial regression coefficients for background markers. Fitting a CIM model may help in two ways, depending whether the markers chosen as cofactors and the target interval are linked. If the included cofactors are not linked to the current region, they may anyway improve the sensitivity of the test, as they likely help in reducing residual variation. If linkage exists, CIM allows to discriminate the target QTL from other linked QTL (Zeng, 1993; Zeng, 1994), even if it is difficult to estimate the joint contribution of multiple linked QTL to the overall genetic variance. The number of CIM cofactors to be included in the model should be carefully chosen. On one side, the use of tightly linked markers may reduce the statistical power when attempting to identify a QTL in a particular region (Zeng *et al.,* 1999). Additionally, the use of too many regressor variables can lead to a bias known as statistical over-fitting (Piepho, 2001). The main limitation with CIM is represented by the one-dimensional nature of its hypothesis testing across the intervals delimitated by flanking markers. Such model seems in fact still inadequate to analyze the genetic architecture of complex traits, which appear to be affected by both multiple QTL and their interactions. To address CIM limitations, Kao *et al*., (1999) introduced a multiple interval mapping (MIM) method for simultaneously fitting by maximum likelihood multiple QTL in the model. MIM is considered to be a more precise and powerful tool for deciphering the genetic bases of complex traits as it reduces the model residual variation. Fitting multiple QTL has also an important role in the estimation of epistasis. With MIM, the more computationally-demanding step is generally represented by the identification of the best-fitting genetic model (number and positions of QTL and epistasis of QTL) in the parameter space.

Significance thresholds in QTL mapping

Applying appropriate significance thresholds in QTL analysis is not trivial, as many factors may influence the distribution of the test statistics, e.g. sample and genome sizes, marker density, number of segregating QTL (Churchill and Doerge, 1994). Several methods have been proposed to set an appropriate level of statistical significance when mapping QTL. Lander and Botstein (1989) relied on the cumulative distribution of the LOD score ("LOD drop-off method") while Churchill and Doerge (1994) estimated empirically the threshold for declaring a QTL, based on permutation tests. Bootstrap-based methods have also been proposed (Efron, 1979; Mammen, 1993).

Experimental populations for linkage mapping of QTL

The creation of an *ad hoc* experimental cross genetically segregating for the trait of interest is the first step in QTL mapping studies and represent an important factor for its success (Darvasi and Soller, 1995). Conventional linkage mapping methods involve the development of segregating population derived from two lines with contrasting phenotype at target trait, with the general aim of being able to observe segregation both at genetic loci and at phenotypic level. Different types of experimental crosses have been proposed (for a review, see Semagn, 2010).

The first experimental crosses which have been utilized for genetic analysis in crops were F_2 or backcross (BC) populations developed from inbred lines in relatively very short time. As a consequence of the presence of long-distance genetic linkage, strong marker-trait associations are possible to be detected even with markers far from the target quantitative locus. The other major limitation of F_2 and BC populations is represented by the fact that they are temporary genetic resources based on single plants and they do not allow replicated trials of the same genotype, except for the cases when cloning (e.g. vegetative propagation) is possible.

Conversely, Recombinant Inbred Line (RIL) represents an immortalized genetic resource derived from an F_2 population through several (generally from five to ten) cycles of selfing in autogamous species (via bulking or single seed descent) or full-sib mating in outcrossing species (Darvasi and Soller, 1995; Soller and Beckman, 1990; Xu and Crouch, 2008). Such multiple generations of selfing/ full-sib mating produce advanced homozygous lines and increase the potential number of recombination events. Due to the breakage of moderate genetic linkage among markers, RIL populations likely allow an higher mapping resolution than previous designs. A shortest way to obtain inbred lines population is the double haploid (DH) strategy, as in just one generation it allows to get complete homozygous individuals at all genomic loci. (*e.g.*, Bao *et al.*, 2002; Xu and Crouch, 2008). A variety of protocols exists in order to produce haploids in different species (e.g. anther/pollen cultures, interspecific crossing). The chromosome

number of haploid plants is generally doubled with colchicine treatment, a substance which inhibits microtubule polymerization in mitosis and permit to develop a DH genetic resource. Another type of permanent population very valuable in genetic mapping of target traits is represented by near isogenic lines (NILs) and introgression lines (IL). Such genetic stocks are characterized by an isogenic background with the exception of the region of the QTL under investigation and could therefore drive to an hypothesis-driven high-resolution mapping. They are usually generated though back-crossing to a recurrent parent at least for six generations, in order to be able to selectively analyse the phenotypic effect attributable to a QTL (Pumphrey *et al*., 2007; Xu and Crouch, 2008). One of the constraints which are common to all the herein described segregating populations for QTL mapping is that the confidence intervals for many detected QTL will correspond to several centimorgans and several hundreds of genes (Doerge, 2002; Holland, 2007).

Apart from increasing population size with the aim of being able to detect more recombination events, mapping resolution can also be improved by providing additional opportunities for effective meiosis during cross development, as in the case of advanced intercross lines (AILs, Darvasi and Soller, 1995). AILs are similar to RILs as they derive from an initial bi-parental cross but involve further cycles of random intercrossing before allele fixation through selfing. Each generation of intercossing likely reduces the extent of linkage disequilibrium, allowing QTL to be mapped more precisely (Rockman and Kruglyak, 2008). This approach has been applied to plant genetics with the name of intermated recombinant inbred lines (iRILs, Winkler *et al*., 2003). An example of iRIL mapping resource is the IBM population in maize (Lee *et al*., 2002), developed by randomly intermating for four generations an F_2 population derived from the inbreds B73 and Mo17. IBM was reported to show a four-fold increased recombination frequencies as compared to an F2 population in maize (Lee *et al*., 2002) and was later used for precise QTL mapping (Rodriguez *et al*., 2008).

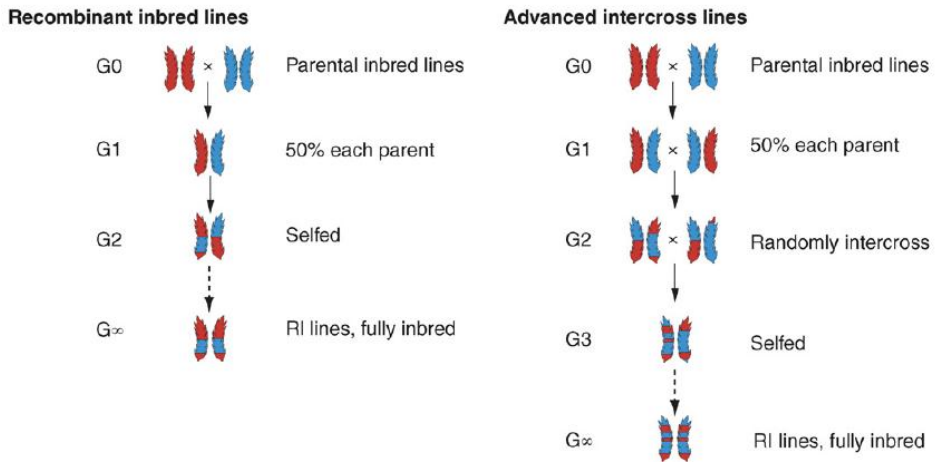


Fig. 3 RIL (on the left) and AIL (on the right) population development (from Cavanagh *et al*., 2008, page 216). The random intercrossing steps pursued with AILs is likely helping in reducing the extent of linkage disequilibrium.

Although they can allow for a more accurate QTL localization, the described multi-meiosis resources are still developed starting from only a pair of founder lines. Indeed, their narrow genetic base cannot significantly represent the variation present in the breeding-relevant gene pool. As pointed out by (Xu, 1996) bi-parental mapping populations are characterized by a statistical inference space limited to the two parents and the QTL analysis is highly population-specific. In order to overcome the intrinsic limitations of conventional methods for mapping QTL in crops, there has been a need to move toward new mapping resources with a broader genetic and phenotypic bases (Blanc *et al*., 2006; Rakshit *et al*., 2012).

2.2 Association mapping

Association mapping (AM) of complex traits is considered as a complementary strategy to conventional QTL linkage mapping based on experimental crosses (see Fig. 4b from Zhu et al., 2008). AM aims to genetically dissect complex phenotypes by exploiting the pattern of linkage disequilibrium (LD, i.e. non-random association of alleles, Flint-Garcia *et al*., 2003) existing in collections of diverse germplasm (Buckler and Thornsberry, 2002; Breseghello and Sorrells, 2006, Yu and Buckler, 2006). AM has been originally developed in human genetics as a promising technique to detect complex disease susceptibility alleles otherwise un-detected in pedigree-based linkage studies (Risch and Merikangas, 1996). In crop genetics, the potential of LD mapping approaches for deciphering the genetic bases of quantitative traits has already been demonstrated (Thornsberry *et al.*, 2001; Gupta *et al.* 2005; Rafalski, 2010). With association

studies, all the recombinations occurred during the history of a sampled population are capitalized in order to identify marker-phenotype statistical associations (Kruglyak, 1999; Jannink *et al*., 2001).

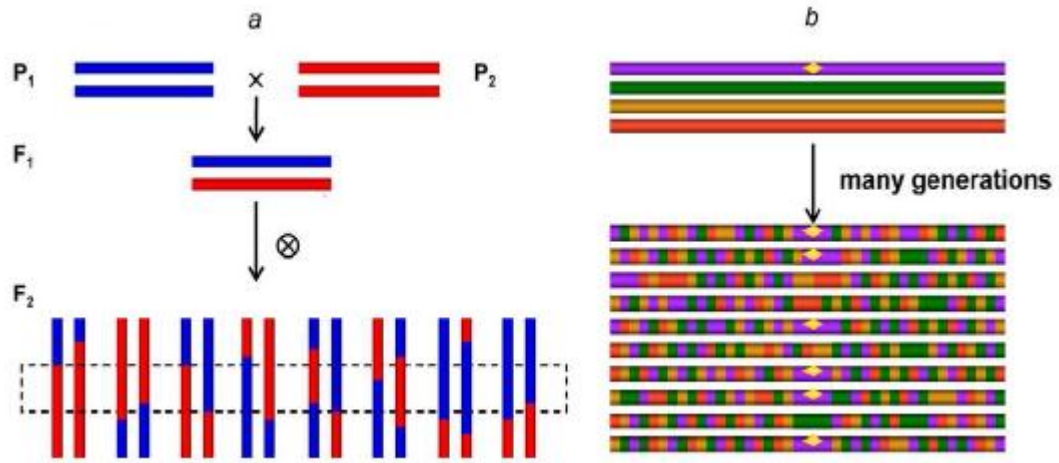


Fig. 4a. QTL mapping approach based on genetic linkage within a bi-parental mapping population. **4b.** Association mapping strategy based on historical linkage disequilibrium existing in collections of diverse germplasm. From: Zhu *et al.,* 2008 (page 6).

Theoretically, AM could achieve an higher mapping resolution in comparison to QTL mapping with bi-parental crosses, since the latter can only rely on the informative meiosis accumulated during cross development (Remington *et al*., 2001; Thornsberry *et al*., 2001; Morgante and Salamini, 2003). However, as proved by Remington *et al*. (2001), the effective AM resolution and detection power is linearly dependent on the extent of LD in the genome and in the particular region under investigation. In addition, in many crop species the genetic bottleneck occurred at domestication or caused by breeding methods caused extensive identity-by–descent of chromosomes among individuals and reduced genetic diversity. These processes have the effect of reducing the informativeness of meiosis and therefore the potential genetic resolution of association mapping. A further advantage in AM compared to bi-parental linkage mapping is the possibility to evaluate a broader spectra of genetic diversity instead of a maximum of two alleles per genomic locus (Buckler and Thornsberry, 2002; Flint-Garcia *et al*., 2003).

In association studies, two kind of approaches could be chosen: whole genome scans (after having saturated the genome with well-dispersed markers to reach a coverage adequate to the estimated LD extent) or candidate gene studies which focused on a particular genomic region with high LD decay to likely identify polymorphisms responsible for variation at trait of interest (Rafalski, 2002b; Thornsberry *et al*., 2001).

The main drawback with association mapping is represented by the high risk of incurring in type I error (false positives) caused by unaccounted subdivisions in the sampled population which can lead to spurious associations (Flint-Garcia *et al,.* 2003; Gupta *et al.,* 2005).

Most germplasm sets are characterized by unknown population structure, i.e. when different subpopulations show different alleles frequencies, as a consequence of their breeding and evolutionary history (Jannink and Walsh, 2002). Factors such as genetic drift, population admixture, genetic bottlenecks may influence LD between markers and quantitative loci, as far as in experimental populations LD appears influenced only by recombination frequency in the absence of segregation distortion (Nordborg and Tavaré, 2002). Association mapping could therefore be hindered by inferential problems caused by both the presence of major sub-populations in the sample under examination and different degrees of relatedness among the lines. A variety of methods have been proposed to infer population substructure in AM studies. They are generally based on clustering individuals either according to distance-based estimates between lines (Nei, 1972) or probabilistically (Pritchard *et al.,* 2000). The most known likelihood-based approach is the bayesian modelling method of Pritchard *et al.,* (2000) embedded in STRUCTURE software (Pritchard *et al.,* 2000), which produces Q (population structure) and K (relative kinship) matrix estimates to be included by unified mixed models in AM studies (Yu *et al*., 2006). An inherent issue is represented by the common removal of rare alleles which may also increase the probability of type I error by inflating LD estimates between unlinked markers (Somers *et al*., 2007).

A<u>ccounting for multiple testing in GWAS</u>

Due to the advances in high-throughput genotyping technologies, the multitude of comparisons performed in a genome-wide association study could easily lead to type I error. An adjustment of statistical test is therefore required. On the other hand, if the adjustment for type I error turns to be overly conservative, a decrease in statistical power may produce false negative results. Such inflation of type II error is more likely to occur with rare variants: functional alleles which appear rare population-wide and for which a decrease in AM statistical power have been described, even if they presented a relatively large effect. A commonly used method aimed at controlling the GWAS-wide Type I error rate is the Bonferroni method, which directly adjust the *p*-value threshold according to the number of multiple comparisons effectuated, assuming that all comparisons are independent. Due to the preponderance of linkage disequilibrium between tested genetic markers, Bonferroni assumption is often violated, leading to type II error (over correction). Several efforts to develop alternatives to Bonferroni correction have been made, in order to better reflecting the dependent nature of genetic datasets. An alternative is offered by

permutation testing (e.g. PRESTO software, Browning, 2008). Another kind of approaches aims to identify the actual number of independent statistical comparisons which are being made (apart from the number of SNP which have been tested). The number of actual comparisons could be defined by Principal components analysis (PCA) methods (Cheverud, 2001; Nyholt, 2004) or estimating the number of LD blocks across the genome (Patterson *et al.,* 2004).

2.3 Multiparental populations

Quantitative Trait Loci (QTL) analysis on multiparental populations (MPPs) occupies an intermediate niche between conventional linkage mapping based on bi-parental populations and association mapping based on collections of germplasm. Linkage mapping with bi-parental resources often detects QTL with large support intervals because of the limited chances for recombination events occurring during population development (Doerge, 2002; Holland, 2007; Li *et al.,* 2010). Allele fixation at tested loci could even determine the impossibility of detecting QTL affecting a certain trait (Xu, 1998).

On the other hand, association mapping approaches (Risch and Merikangas, 1996; Nordborg and Tavaré, 2002) relying on both wider genetic and phenotypic diversity and historical recombinations are often hindered by inferential problems caused by hidden population structure (Balding, 2006).

The idea of combining the genomes of multiple founders in a single multi-meiosis mapping population to ensure segregation for multiple QTL for multiple traits was first developed in the framework of mouse genetics with the name of heterogeneous stocks (HS, McClearn, 1970; Talbot *et al,.* 1999; Demarest *et al.,* 2001; Valdar *et al*., 2003). HS could be considered as an extension of the bi-parental AIL approach, involving multiple parents in the crossing scheme in order to include a broader genetic diversity. During HS development, eight mice founder strains have been intercrossed for more than sixty generations (see Mott *et al*., 2000). Each HS line chromosome can be considered as a fine mosaic of the known founder haplotypes and specific software have been created to reconstruct the probable ancestry of each genomic segment (Mott *et al*., 2000). To prevent fixation of alleles by genetic drift, a large population of mating pairs (forty lines) has been provided at each generation. RI lines can be then derived from an heterogeneous stock by repeated brother-sister mating. Yalchin *et al*. (2005) they have successfully used HS resource to fine QTL controlling complex traits in mice to extremely narrow confidence intervals. To further address complex trait dissection, mouse-genetics community (the Complex Trait Consortium) has also established a large multiparent RIL panel

derived from eight laboratory mouse inbred strains which was called the Collaborative Cross (CC, Churchill *et al.,* 2004 Threadgill *et al*., 2011).

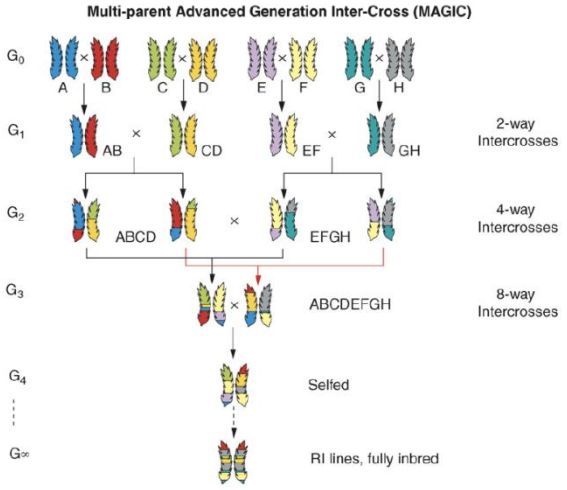


Fig. 5 Development of a MAGIC genetic resource (from Cavanagh *et al.,* 2008, page 216).

A multiparent strategy was then adapted to plant genetics via multi-parent advanced generation inter-crosses (MAGIC, Mackay and Powell, 2007; Cavanagh *et al*., 2008) and interconnected populations via di-allelic schemes or star designs (Yu *et al*., 2008; Huang *et al*., 2011). A typical MAGIC resource is initiated by a funnel breeding scheme in order to combine the founder genomes; the resulting intermediate lines are then randomly intercrossed for a varying number of generations prior to fixation (see Fig. 5). Such inter-crossing steps are pursued in order to shuffle the founder genomes to the point that each fixed recombinant inbred line (RIL) can be seen as a particular mosaic of the initial lines.

Mapping of QTL with MPPs indeed combines a greater accuracy due to the high number of informative crossovers which cause linkage disequilibrium breakdown (Broman, 2005; Flint *et al*., 2005) and the possibility to interrogate multiple alleles over multiple genetic backgrounds with a single study (Rebai and Goffinet, 2000).

With multiparental populations, both linkage and association methodologies can be conducted, without generally being hampered by population structure (Brachi *et al*., 2010).

Current software tools for linkage mapping in MAGIC population rely on a variety of statistic methods for reconstructing the parental haplotype probabilities along the line fixed chromosomes and being able to infer the parental origin of allelic information (Broman, 2005; Teuscher and Broman, 2007, see section 5.3). In connected bi-parental populations allelic ancestry information is not ambiguous but it has been proved that fitting a joint multi-

population model with IBD estimates (rather than nesting effects whithin populations) could increases the power of QTL detection (Rebai and Goffinet, 1993; Jannink and Jansen, 2001) and improve the accuracy in estimating the location and allelic effect of detected loci (Meuwissen and Goddard, 2001; Jansen *et al*., 2003; Uleberg and Meuwissen, 2007, see section 5.3).

If enough RILs are generated, MPPs offer as well the possibility to examine epistasis (Charcosset *et al.*, 1994; Rebai *et al*., 1994; Jannink and Jansen, 2001) and estimate GxE interactions to model more realistically the genetic architecture associated with complex traits. As pointed by Charcosset (Charcosset *et al.,* 1994) connected multiparent design enables to investigate higher-order epistasis than digenic interactions.

These theoretical advantages of MPPs are now being tested empirically with existing populations and tools. Since preparing a multiparent genetic resource is demanding in terms of time, labour and costs (Rakshit *et al.,* 2012), few have been created, and each one in a new species provides a valuable platform for a wide variety of analyses. In plants, different types of MPPs have been produced and utilized for mapping QTL: *Arabidopsis thaliana* MAGIC and AMPRIL populations (Kover *et al*., 2009; Huang *et al*., 2011); four and eight-way panels in bread wheat (Huang *et al*., 2012); maize NAM design (Yu *et al*., 2008); Arabidopsis NAM design (Bentsink *et al*., 2010; Brachi *et al*., 2010) and four MPPs in rice (Bandillo *et al*., 2013).

3. Molecular-based genetic improvement (from MAS to GS)

Following the advancement of molecular marker technology, marker-assisted selection (MAS) has gained a fundamental role in breeding programs targeting quantitative complex traits of agronomic relevance. Molecular-based crop genetic improvement has in many cases substituted conventional selection practices which merely relied on phenotypic evaluation (not a reliable predictor of individual breeding values, especially with traits characterized by a low hereditability and a complex genetic background). Several quantitative loci are often detected as associated with a complex phenotype and, as a consequence of the high LD encountered in experimental crosses, markers can reliable act as predicting variables (Whittaker *et al*., 1995). MAS approach involves the early-stage marker-based identification of those individuals showing the favourable alleles of interest at detected QTL, whose population frequency is then generally increased with several breeding cycles, e.g**.** through marker assisted recurrent selection (MARS) strategies (Bernardo *et al*., 2006, Lorenzana and Bernardo, 2009).

Predicting the probability of allele transmission in different MAS schemes is not straightforward, decision support platform in marker assisted breeding (see OptiMAS, Valente *et al*., 2013) may greatly help in marker-assisted assembly of diverse alleles of agronomic values in new genetic materials. Some authors (e.g. Hospital *et al*., 1997) underlined that the possible fixation of unfavourable alleles at loci with small effect on the phenotype (often undetected in QTL mapping) could significantly affect the efficiency of MAS over breeding cycles. Similarly, after the first generation, marker tagging could not hold in the long term as a result of unstable marker-phenotype associations. Regarding the first issue, Hospital *et al.* (2000) have stressed on the importance of equally weighting all index predictor loci, whatever the magnitude of the recorded effect. To address the stability of MAS predictions molecular breeding has been recently implemented with experimental population which could represent a larger breeding pool. However, the well-known constraints related to the statistical power in QTL detection experiments appear as the main limiting factor of MAS practices (Beavis, 1994, Hospital *et al.,* 1997).

Whittaker *et al.* (2000) proposed though simulations a shift in the paradigms of quantitative genetics related to breeding improvement: avoiding the selection of significant markers by using a penalized regression which involved all genotyped loci.

In the context of animal breeding, Meuwissen *et al.* (2001) suggested in fact to skip the step involving the detection of significant marker-trait associations in favour of a genome-wide approach based on SNPs as unit of selection. The rationale of such approach, which has been called genomic selection (GS), is to take advantage from dense marker data which is nowadays from high-throughput genotyping platforms and base the selection step on aggregates of estimated marker effects, summarized by predicted genomic breeding values (GEBV). The GS model will therefore likely capture all genomic variation associated with a certain phenotypic performance, including also the effect of minor QTL too small to be declared as significant in QTL mapping experiments. In practice, predictive models are generally trained only on a representative subset of individuals (called training population) for which both phenotypic and genotypic data have been recorded (Schefers and Weigel, 2012).Two major issue have to be faced when simultaneously estimating genome-wide marker effects. First of all, due to huge density of marker data, more predictor (p) than observations (n) is likely to arise when fitting the model with least square methods, meaning that the degree of freedom of the model are not sufficient to simultaneously estimate all predictor effects (the so-called 'large p small N problem'). The genome-wide model trained on the reference set is then applied to the entire set of individuals (the selection candidates having only genotype data) in order to predict GEBV and select the candidates for advancement in the breeding cycle. Several studies addressed the

selection efficiency of GS both in animal and plant breeding and highlighted a significant improvement with regards to conventional MAS based on BLUP models and QTL detection (Hayes *et al.*, 2009; Heffner *et al*., 2011). Genome-wide prediction accuracy was also assessed analysing diversity panels and multiparental resources, which exploit a broad genetic base with the aim of identifying favourable alleles (Crossa *et al*., 2010, Rincent *et al*., 2012; Bardol *et al.,* 2013). Although multiparental designs investigate a more limited sample of diversity then association panels, they generally represent breeding-relevant materials and appear closer to real breeding practices (Würschum, 2012). In animal breeding context, Calus *et al.* (2008) pointed out that the use of haplotype information allow to significantly enhance the prediction accuracy of GS and to better model variation at multi-allelic QTL.

The importance of determining the actual haplotypes segregating in crop breeding-relevant gene pools has been extensively underlined (see the 'breeding by design' concept by Peleman and Rouppe van der Voort, 2003). In the plant breeding framework, a currently debated question is whether more targeted approaches (relying on 'chromosome haplotyping' to identify the allelic variation at target loci, Peleman and Rouppe van der Voort, 2003) are more effective than GS in selecting optimal combination of alleles (Lübberstedt, 2013). Efforts in the fine-scale definition of the pedigree haplotypes (ancestral chromosome blocks) present in breeding-relevant materials have been recently carried out in rice (Yamamoto *et al*., 2010). From a breeding point of view, describing the association between particular haplotypes and phenotypic variation in crops could represent a useful tool to allow breeders to avoid redundant haplotypes in crossing designs and facilitate future targeted selection (Yamamoto *et al*., 2010).

4. Single nucleotide polymorphisms

SNPs (i.e. single nucleotide polymorphisms) represent the richest source of sequence genomic variation in populations (Brookes, 1999). Due to their high abundance in virtually all populations of individuals and relatively low mutation rate they can be easily converted into very stable genetic markers suitable to highly multiplexed and relatively low-cost automated analysis (Rafalski, 2002a). Several high-throughput SNP-genotyping platforms are now available for profiling up to one million SNPs in parallel (e.g. Illumina GoldenGate, Illumina Infinium, Affymetrix GeneChip). As a result, SNP technology is nowadays going to be currently preferred to previously widespread marker systems based on Simple Sequence Repeats (SSR) and Diversity Arrays Technology (DArT) markers (Ganal *et al.,* 2009; Varshney *et al*., 2009). SNPs are considered an efficient marker tool in plant species. Their main applications involve the construction of high-resolution genetic maps, the assessing of linkage disequilibrium patterns and genetic diversity, the discovery of marker-trait associations in mapping complex phenotype, marker-assisted breeding (Rafalski *et al*., 2002a) and genomic selection (Meuwissen *et al*., 2001).

4.1 SNP discovery in crops

As expected, heavy SNP discovery efforts are required for the construction of large SNP-genotyping platforms with whole-genome coverage. Several different approaches for identifying and validating SNPs within the genome of a certain species may be adopted (see Edwards *et al*., 2007). The simplest route for SNP discovery in a defined target region, involves the comparative analysis of PCR-amplified sequences generated from different individuals and the scanning for previously unknown polymorphisms. Alternatively, the multitude of published expressed sequence tags (EST) libraries (URL: http://www.ncbi.nlm.nih.gov/dbEST) represents a considerable source for "in silico" SNP detection through bioinformatics pipelines. ESTs are developed by single-run sequencing of cDNAs obtained from different individuals and are usually redundant in databases. Thus, overlapping sequences are assembled in multiple alignments and mismatches are reported as candidate SNPs. Pipelines generally comprise the estimation of a base calling quality score (as Phred, Ewing *et al*., 1998a and 1998b) to reduce the likelihood of false positives when calling a putative SNP without a previous step of *de novo* sequencing. Whole-genome sequencing projects offers unprecedented opportunities for identifying single nucleotide variation among individuals. Comparison of overlapping BAC clone sequences could also provide opportunities for identifying novel polymorphisms. The number of sequenced plant genomes is constantly increasing but efficient whole-genome re-sequencing

SNP discovery have been reported only for genomes with low-to moderate complexity as *Arabidopsis*, rice, barley and maize (Kumar *et al*., 2012b). As a result, the set of SNPs available for un-sequenced and complex sequenced genomes is still low, and such issue delayed the utilization of SNP platforms in crops (Ganal *et al*., 2009). In recent years, technological advances such as next-generation sequencing (NGS, Metzker, 2010) has enabled faster rates of *de novo* and reference-based SNP discovery for numerous plant species. With NGS, a SNP is identified after comparing reads from different genotypes (or comparing reads to the reference genome) and adjusting for the genome coverage achieved by the NGS experiment. Moreover, as large parts of plant genomes consist of repetitive element, specific algorithms for enabling correct SNP assembling are particularly important to prevent erroneous read mapping to paralogous sequences (Kumar *et al*., 2012b). Similar misalignment could potentially be created by homoeologous regions in polyploid genomes (see section below). In such cases, the identification of SNPs by parallel sequencing technology has been combined with reduced complexity approaches aimed to reduce the representation of low-information-content repetitive sequences (e.g. by generating reduced representation libraries (RRLs) or by CRoPS technology).

4.2 The problem of SNP identification and use in wheat

Bread and durum wheat are allopoplyploid crops characterized by 21 pairs of homologous chromosomes and 14 pairs of homologous chromosomes, respectively, each composed of 7 homoeologous groups (see section 1.1). The main hurdles to a large-scale SNP identification in wheat are the low level of genetic diversity encountered in the breeding germplasm, the absence of a whole-genome reference sequence, and the highly repetitive and the duplicated nature of the genome (Somers *et al.,* 2003; Ganal and Röder, 2007; Barker and Edwards 2009; Ganal *et al.,* 2009). The low nucleotide diversity which characterizes bread and durum wheat has been likely caused by the severe species formation (i.e. ploidy change) and domestication bottlenecks experienced by the cultivated wheat species (Ravel *et al*., 2006; Haudry *et al*., 2007; Ganal *et al.,* 2009). Moreover, a high number of paralogous genes have been described to characterize wheat genome (Dubcovsky and Dvorak, 2007) and the percentage of repetitiveness has been estimated to be approximately equal to 77% (Flavell *et al*., 1977). Such issues reduces the efficiency of both SNP discovery and genotyping, as closely related paralogues, pseudogenes and, generally, multi-copy sequences could confound bot SNP discovery and the downstream ability of correctly identifying the allelic state of each individual at marker loci. (Akhunov *et al*., 2009).

An additional concern when working with SNPs in wheat species is the possible confounding role of the homoeologous genome/s, which were estimated to share a sequence identity of around 96-98% (Dvorak *et al*., 2006). Therefore, homoeologous sequence variants have a considerable chance to act as confounding factors when analysing SNPs. As summarized by Trick *et al.* (Trick *et al*., 2009), polyploid genomes are characterized by three types of polymorphism (see Table 1), namely 1) inter-homoeologue polymorphisms 2) inter-varietal SNPs and 3) hemi-SNPs.

Homoeo-SNPs (e.g. in tetraploid durum wheat, AA/AB vs AA/AA or BB/BA vs BB/BB), i.e inter-homoeologue polymorphisms or "false SNPs", represent the most frequent class of polymorphism occurring within two homozygous individuals (Somers *et al*., 2003; Ravel *et al*., 2006; Barker and Edwards, 2009). As homoeo-SNPs are not expressing allelic variation *per se* and are possible to detect also paralogous loci within each of the genomes, most of the genotypes called with an homoeo-SNP marker will result in heterologous loci with lesser value for mapping purpose.

By contrast, simple SNPs (i.e. inter-varietal SNPs or true SNPs) are derived from allelic differences at a single genomic locus. Being a genome-specific assay they could thus be assimilated to a simple diploid tests. This SNP class reflects varietal SNPs between individuals and is traditionally referred as allelic variation, essential for mapping purposes. Due to selection pressure, simple SNPs represent only a very limited proportion (10-30%) of total polymorphic SNPs in various polyploidy crop species (Ravel *et al*., 2006; Barker and Edwards, 2009; Mammadov *et al*., 2012).

The major class of hemi-SNPs (e.g. in durum wheat: AABB vs BBBB or AABB vs AAAA) represents a SNP assay which is amplifying both genomes but is homozygous in one genome and heterozygous in the other. It is therefore considered as an allelic variant observed in the presence of homoeologous sequences, scored as dominant marker as the most frequent base is un-informative. As reported by Mammadov *et al*. (2012), hemi-SNPs represent a percentage varying from 30 to 60% of SNP variation in polyploidy crop species.

Table 1 SNP classes encountered in polyploid wheat genomes. Cv1 and Cv2 indicates the putative genotypes of two different cultivars.

	Genome A	Genome B	
Cv1	11	10	Homoeo-SNPs
Cv2	11	11	
Cv1	11	- -	Simple SNPs
Cv2	0	- -	
Cv1	11	00	Hemi-SNPs
Cv2	00	00	

5. Haplotype analysis

An haplotype is defined as a specific combination of alleles occurring (cis) on the same chromosome. The human genome has been described as comprising small regions of high recombination frequency (recombination hotspots) connected by conserved segments characterized by strong linkage disequilibrium between markers (Daly *et al*., 2001; Patil *et al.,* 2001; Gabriel *et al.,* 2002). Such SNP association patterns have been termed haplotype blocks and have been inherited from generation to generation essentially as a single unit where only limited recombination has occurred since the origin of modern humans (Daly *et al*., 2001; Jeffreys *et al.,* 2001). Such structural arrangement holds for cereal species too (Comadran *et al*., 2011; Cavanagh *et al*., 2013).

It has been hypothesized (Risch and Merikangas, 1996; Pritchard, 2001) that the concurrence of multiple susceptibility alleles of independent origin could underline many complex diseases in populations. Haplotype-based association studies are currently considered as more effective (in terms of statistical power) than individual SNPs for investigating quantitative loci (Slager *et al*., 2000; Longmate, 2001). In fact, in presence of multiple susceptibility alleles (and particularly when linkage disequilibrium between SNPs in a target region is weak) association analysis based on haplotype profiles determined after SNP genotyping can be more informative over the use of individual SNP binary data when an allelic series exists at a locus (Morris and Kaplan, 2002). Apart from depicting a more discriminative state of a chromosomal region between case and controls in disease association studies, the use of haplotypes also reduces the number of tests to be carried out and hence the penalty for multiple testing (Zhao *et al*., 2007). The advantages of an haplotype-based analysis depend on various factors as: the trait genetic architecture, the marker density and the local LD pattern (Lorenz *et al.*, 2010). In 2002, an international collaboration (called the International HapMap Consortium) has been set up with the goal of developing an haplotype map of the human genome (HapMap), which will describe the common patterns of genetic variation in humans.

Haplotype definition and haplotype-based QTL mapping are receiving an increasing interest also in the framework of complex trait mapping in crop species. As herein described in human genetic context, considering multi-SNP haplotypes as a synthetic multi-allelic marker system could combine SNP ubiquity and abundance with the advantages of an haplotype-based analysis: low redundancy, high informativity and power in analysing association with phenotypes. Haplotype data could in fact capture associations that would elude analysis based on single SNPs, e.g. when the number of QTL functional alleles could not being fully described by the distribution of marker variants. Haplotype analysis in crops offers two additional interesting properties. The analysis of the haplotype structure in domesticated plants could put new

highlights into the historical flow of pedigree haplotypes during breeding practices, tagging the genomic regions which have been subjected to more intense selection (Yamamoto *et al*., 2010). From an applicative point of view, a QTL analysis based on haplotypes could allow haplotype-assisted selection and, more generally, haplotypes and IBD/IBS investigations at target loci to enable a more informed classification of parental lines.

The computational issues related to haplotype analysis could be subdivided in two major conceptual subcategories. On one side, the set of methods for inferring the allelic state of a certain chromatid on precise individual (i.e. haplotype phasing). Regarding plant genetics, phasing is not an important issue in crop species (such as wheats) which are commonly handled as inbred lines. Yet, other crop species (e.g. many fruit crops) suffer of the phasing problem. Additionally, a problem similar to the haplotype phasing must be dealt in recent polyploid species with similar homologous genome sets such as wheats. Moving to the population-level, haplotype reconstruction refers to a comprehensive set of methods which have been developed in order to locally define haplotypes along the genome and locally infer IBS/IBD states.

5.1 Haplotype inference

In diploid organisms including human, current high-throughput genotyping techniques return information about which alleles are present at each typed SNP but are unable to discriminate data coming from two homologous chromosomes and therefore loose haplotype phases (i.e. the gametic phase results ambiguous). Knowing the haplotypes at individual's parents could help in uniquely define phases at target region only if pedigree data is fully informative. Due to the increased interest in haplotypes, a variety of computational methods have been proposed to infer phase at linked loci from individual genotype data when parental genotypes are ambiguous (for a review, see Browning and Browning, 2011). Generically, an haplotype resolution problem is solved using probabilistic models based on observations carried out in a sample of individuals. A well-known computational method phasing haplotypes is Clark's parsimony approach (Clark, 1990) which attempt to minimize the total number of haplotypes observed in the sample with a combinatorial approach starting from those individuals whose haplotypes are unambiguously inferred from their genotypes (i.e. individuals homozygous at every locus or heterozygous at only one locus). Other widespread inferential methods rely on likelihood functions combined with optimization algorithms (e.g. E-M, MCMC, HMM) based on different assumptions, e.g. the limited diversity of the haplotypes in the population (Excoffier and Slatkin, 1995; Long *et al*., 1995; Fallin and Schork, 2000). In particular, a wide-spread assumption is the *infinite sites model* by Hudson (1990) and the correlated perfect phylogeny model (Gusfield, 2002): at each SNP site

a mutations occurred only once in the history of a population, tagging all the possible descendants of a single ancestors. Other examples of statistical methods for haplotype reconstruction are the Gibb sampler in PHASE (Stephens *et al*., 2001) and HAPLOTYPER (Niu *et al*., 2002).

5.2 The definition of haplotype blocks

The definition of haplotypes at a particular region strictly depend on the length of the genomic window being examined: the larger the window, more haplotype classes are likely to be expected in a particular sample of individuals as more chances for recombination events have occurred. In regions characterized by strong LD, many SNPs comprised in a particular haplotype contain redundant information and a small subset of them (called tagging SNPs) may suffice to completely define local haplotypes. An example of statistical software able to infer haplotype patterns and tagging SNPs from genotype data is Haploview (Barrett *et al*., 2005). Haplotype block definition with Haploview is performed either by LD structure estimation (Confidence interval or Solid Spine methods) or by the alternative four-gametes rule (Wang *et al*., 2002) which compute the population frequencies of the 4 possible two-marker haplotypes for each marker pair to detect recombination events.

5.3 Haplotype reconstruction in MPPs in crops

Haplotype reconstruction is a pivotal issue when analysing multiparental crosses in crops. The step of defining the parental origin of marker alleles along the fixed chromosomes have to precede any genetic analysis and statistic inference in any multi-way population. Then, computational tools for multiparental linkage map construction, missing data imputation and mapping of quantitative traits must be able to contend with all the allelic segregation patterns consistent with observed marker genotypes and accommodate the framework which is better fitting the associated founder probabilities.

Although marker phases are generally known due to the inbred state of parents, the ambiguous inference regarding the parental origin of allelic information is generally due to two major causes. The first one is correlated to the information provided by the generally adopted marker systems. Due to the high-throughput of SNP platforms, SNPs are often the marker system of choice. While such co-dominant bi-allelic marker could be either fully informative (each genotype probability is uniquely identified) or non-informative in a simple cross, a SNP marker system will never be totally informative in a multi-parent population. In a MPP, even if up to n

(n=number of founders) alleles can theoretically segregate at each tested genomic locus, a SNP marker system will capture no more than two alleles in each investigated position. It is therefore impossible to discriminate alleles that are identical by state from those that are identical by descent without applying probabilistic models accounting for founder haplotype probabilities along the genome. The second major cause which hampers the possibility of inferring the ancestry of any locus on crop MPP's genomes is represented by the severe bottlenecks which characterized the history of cultivated crops. As a result, many genomic regions are expected to show IBD stretches, impeding direct observation of the parental origin of marker alleles as the number of alleles segregating in the final population is often inferior to the actual number of founder lines. In such cases, choosing an hypothetical multi-allelic marker system instead of wide-spread SNP platforms could not enhance the informativity of the test.

The theoretical framework for haplotype reconstruction in MPPs have been already extensively defined (Broman, 2005; Teuscher and Broman, 2007). Nowadays, different statistical tools are available for reconstructing the genome of each line as a mosaic of the founder haplotypes and trace back to the founder lines the present allelic information.

Some of these computational approaches (e.g. R/Happy, Mott *et al*., 2000; Valdar *et al*., 2006) ignore pedigree information and initially consider all ancestry combination as equally possible. Then, based on HMM (where the hidden states are the progenitor haplotypes and the observed data are the final line genotypes, Broman and Sen, 2009), founder assignment is inferred along the individual chromosomes. A more accurate inference could be usually performed by methods accounting for the breeding design, e.g. including the three-point probabilities computed conditionally on the observed marker data at flanking markers described in Broman (Broman, 2005) or multipoint probabilities-given all the linked-marker data (as provided in R/qtl, Broman et al., 2003). Huang and George (2011) developed a R-coded platform for analysing MAGIC populations with the possibility or interfacing both to R/qtl and R/Happy to infer founder probabilities at each locus.

Haplotype definition has also been adopted for mapping QTL in multiple connected bi-parental populations. Statistical models used for mapping QTL in such MPP designs are generally based on the assumption of parental allele independency (i.e. the existence of a different QTL allele for each segregating parental allele). The independency hypothesis generally does not hold in crops populations, as the parental lines are often related and at a given tested position they may share identical alleles from common ancestors.

In the context of animal genetics, Meuwissen and Goddard (2001) proposed to analyse multiple connected designs with an alternative approach called Linkage Disequilibrium-Linkage Analysis (LDLA). It combines conventional QTL mapping based on the genetic linkage (deriving from recent recombination within families) with estimated identity by decent (IBD) probabilities between founders. Inspired from methods developed in human and animal genetics, Jansen *et al.* (2003) proposed an advanced haplotype-based method for QTL mapping of multi-allelic plant populations where parental alleles are grouped based on local marker similarity of parents.

LDLA method generally involves the computation of IBD probabilities from dense marker coverage of parents to estimate local correlation between random effects of QTL alleles. Such LD information (computed trough pedigree and/or marker information, Van Eeuwijk *et al.*, 2010) is usually stored in a variance-covariance matrix of QTL alleles and fitted in a mixed model for QTL analysis. Many simulation-based as well as empirical studies highlighted the potentialities of LDLA method in comparison to conventional linkage analysis for QTL mapping: a significant gain of power, and accuracy in QTL mapping and allelic effect estimates has been reported (Meuwissen and Goddard, 2001; Jansen *et al*., 2003; Uleberg and Meuwissen, 2007). As pointed by Jansens *et al*. (2003), LDLA promising features are likely due to a statistical issue: the reduction in the number of parameters to be estimated in the QTL mapping model after the clustering of founder alleles into different classes of ancestral alleles. Such putative ancestral classes are in fact modelled with the same parameter, leading to a more parsimonious (and statistically more powerful) model. Marker-based methods for locally clustering parental information in LDLA studies have recently been discussed (Bardol *et al.*, 2013). A bi-allelic single-marker clustering was in fact compared to a more computationally-demanding clustering methods based on similarities between local marker haplotypes (ClustHaplo software, Leroux *et al*., 2014).

OBJECTIVES

Multiparental cross designs for mapping quantitative trait loci in crop species are raising as efficient alternative to conventional experimental populations involving only two founders, since they enable to explore a wider genetic variability and provide a higher mapping resolution.

The present research focused on the analysis of a new multiparental population in durum wheat (*Triticum durum* Def.) named NCCR, derived from a four-way cross involving the cultivars Neodur, Claudio, Colosseo, Rascon/Tarro. Based on pedigree information of parental cultivars, this cross represents well the genetic diversity of elite durum wheat germaplasm. NCCR population is composed by 338 $F_{7:8}$ recombinant inbred lines and segregates for several traits of agronomic relevance, therefore it represents a valuable resource for multi-trait quantitative loci mapping analysis. As a multiparental resource initiated by elite germplasm, NCCR offers also the chance to evaluate the effect of multiple QTL alleles/haplotypes in a breeding-relevant genetic background. The work herein described regarded four main specific objectives:

- Genotypic and phenotypic characterization of NCCR
- Development of a cluster file for correct high-throughput SNP genotype calling in tetraploid wheat
- Construction of a linkage map
- Quantitative trait loci (QTL) analysis for yield, yield-components and other traits of agronomic relevance by both a bi-allelic single marker analysis and an interval mapping analysis based on the founder haplotype probabilities

MATERIALS AND METHODS

1. Plant material and crossing design

A balanced four-way multiparental cross was developed starting from four durum wheat cultivars (Neodur, Claudio, Colosseo and Rascon/Tarro) chosen as diverse contributors of different alleles of agronomic relevance. Neodur (184.7/Valdur/Edmore) is a photoperiod-sensitive late cultivar showing an high number of spikelets per ear; Claudio (Sel. CIMMYT 35/Durango/IS1938/Grazia) shows wide adaptability, stable high test weight and high number of fertile tillers; Colosseo (Creso-derived with Italian landrace introgressions) presents high-yielding ears (in terms of thousands kernel weight); Rascon/Tarro (Rascon-37/2*Tarro-2, CIMMYT germplasm) is a photoperiod-insensitive cultivar with high yield potential due to an high number of grains per spikelet. Indeeed, yield potential of the four selected lines seemed to be driven by different yield components.

Regarding resistant/tolerance to pathogens, the four cultivars presented the following favourable features: Neodur appeared characterized by partial tolerance to *Fusarium* and *SBCMV*; Claudio by resistance to powdery mildew; Colosseo showed brown rust resistance.

Besides showing valuable agronomic traits, the four founders appeared also to well-represent durum wheat elite germplasm, as it is reported in the PCA plot in Fig. 1 (from Maccaferri et al. 2003, pag. 791). Neodur belongs to the 2E group of the US-related cvs; Claudio is represented in the plot by Grazia cv. (group 2A: Valnova-Valforte related cvs); Colosseo appears to be situated in the group 2B (Creso-related cvs) and Altar84 (group 2D: Altar 84-related cvs) acts for Rascon/Tarro in the PCoA.

In the present work, the genetic diversity represented by the four founders has been anew estimated by means of an UPGMA cluster analysis (simple matching genetic similarity, Tassel v.4) based on 280 durum wheat cultivars and accessions, genotyped with a 90k-wheat chip.

The cultivars were pairwise crossed following the scheme (Neodur / Claudio // Colosseo / (Rascon*Tarro), i.e. NCCR) to produce 2-way F1s which were subsequently crossed to produce 400 4-way F1 NCCR hybrids (Fig. 2). These 4-way F1s were advanced through SSD and have been bulked in the F8 generation.

Table 1 Pedigree, origin and year of release of the durum wheat cultivars utilized as founders of the NCCR population.

Cultivar	Pedigree	Origin	Year
Neodur (Neo)	184-7/Valdur//Edmore	France	1987
Claudio (Cla)	CIMMYTselection/Durango//IS193b/Grazia	Italy	1999
Colosseo (Col)	Mexa mutant/Creso	Italy	1995
Rascon/Tarro (RT)	Rascon_37/2*Tarro_2	Mexico	2000

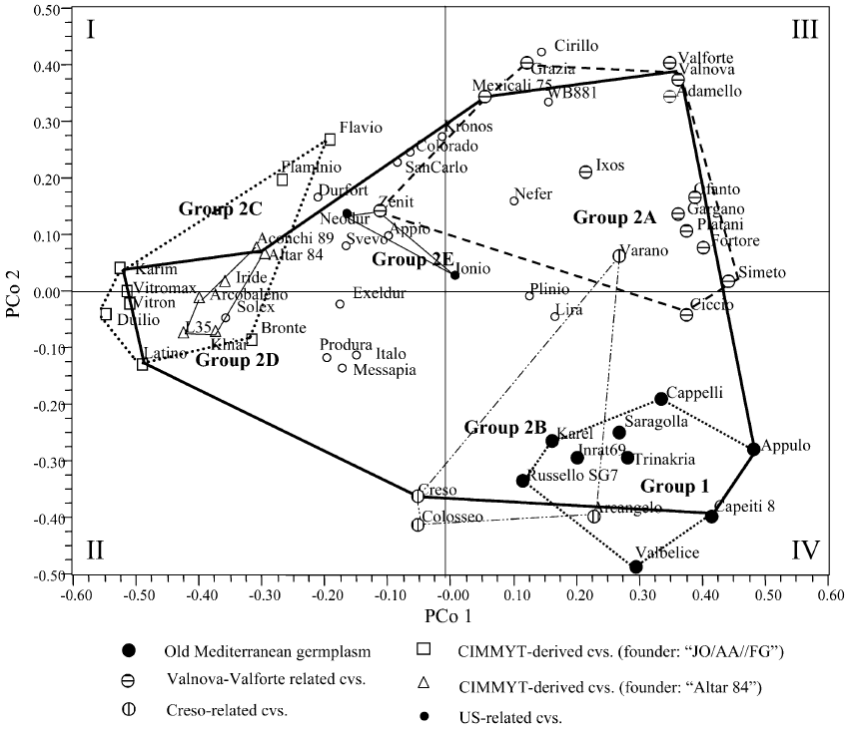


Fig. 1 Principal Coordinate Analysis (PCoA, principal coordinate 1 and 2) on 58 durum wheat accessions as reported in Maccaferri *et al*. 2003, pag. 791. Five main gene pools have been identified (convex hulls highlighted the most diverse cultivar related to each pool).

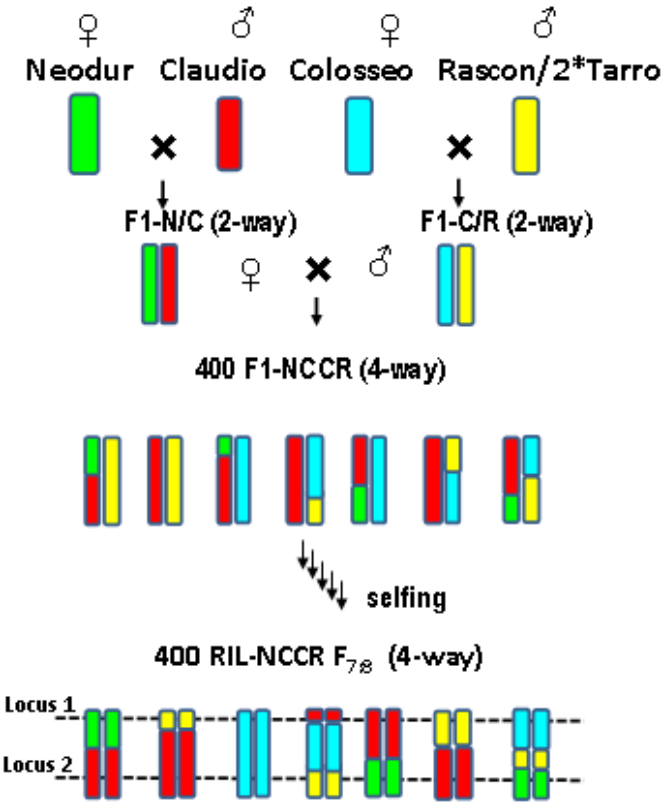


Fig. 2 NCCR four-way RIL population development.

2. NCCR genotyping

2.1 Sample preparation

For each NCCR RIL, we sampled leaf tissue from four plants. Plant tissues were then lyophilized and DNA extraction was performed using Qiagen DNeasy kit standard protocol (Qiagen). By means of a Tecan infinite 200Pro plate reader we normalized NCCR plates to 50 ng/µL.

2.2 High-throughput genotyping and SNP chip description

Genotyping was carried out by means of a wheat-dedicated 90K Illumina Infinium array inlcuding 81,587 effective SNPs and recently developed by an international consortium (E. Akhunov, KSU e M. Hayden, Victorian AgBiosciences. (Wang *et al.*, in press).

The array, an Infinium iSelect HD custom BeadChip, allowed to high-throughput genotype NCCR via a DNA hybridization-based technology. Technically, a BeadChip is composed by silica beads of 3 µm in diameter carrying (covalently linked) locus-specific 50-mers, randomly assembled into miniaturized wells.

The Infinium HD technology combines two different strategies for probe assaying, namely Infinium I and II, each of them specific for different SNP typologies. The Infinium I strategy carries one different probe, i.e. beadtype, for each allele. It is recommended with less common A/T and C/G SNP alleles. On the contrary, the Infinium II strategy (suitable for most SNPs) presents one single probe for both alleles and carries at the 3' terminus the base just before the SNP locus under investigation. In both cases (Infinium I and II), each SNP assay is represented in each chip with a 15 to 30-fold redundancy, to increase the likelihood of correct genotype calls.

The Infinium HD protocol requires around 750 ng of genomic DNA for each sample and involves an isothermal whole-genome amplification, followed by DNA digestion with restriction endonucleases (Kennedy *et al*., 2003). Selectivity of the marker assays is accomplished through sample hybridization to immobilized probes (the locus-specific 50-mers) capillary flow-through chamber. Allele specificity is guarantee by high-fidelity allele-specific extension (by SBE, single base extension).

Allele discrimination is pursued during the SBE reaction with the help of a single hapten-labelled di-deoxynucleotides and a later multi-layer immunohistochemical sandwich assay, whole signals are recorded by an Illumina's scan imaging system. The probes characterized by the Infinium II

chemistry, are fluorescently stained with a dual colour channel (Cy5-Red and Cy3-Green dyes representing each allele).

Successively, the raw signal intensities (X and Y) referring to the two possible SNP alleles are experiment-wide normalized and usually processed with a clustering algorithm in order to convert the continuous signal scores to discrete genotype classes. With Illumina BeadChip platforms, the GenCall Illumina's method implemented in the GenomeStudio software is usually adopted.

Two polar coordinates are estimated from the raw intensity data: R and θ. R (R = X + Y) measure the total signal intensity for each marker test and θ (θ = 2/π*arctan(Y/X)) refers to the angle deviation from pure A-signal (genotype AA, θ=0) in a polar intensity plot representing A-normalized intensity on the x-axis and B-normalized intensity on the y-axis. Pure B-signal (genotype BB) is represented by the maximum signal value (θ = 1), so theta it is also called 'BAF' (B allele frequency) as it represents the B-signal intensity ratio.

The Illumina's GenCall method for allele scoring relies on sample clustering is based on the expected position of the homozygote and heterozygote genotypes for each particular SNP in a diploid genetic model (expected R and theta values of the cluster position for every genotype and SNP are embedded in a so-called 'Clusterfile', which is a Cluster *.egt file to be input in the GenCall run).

The Illumina algorithm therefore defines the allowable signal intensity ranges for AA, AB, and BB samples and, eventually, it estimates the predicted intensities of missing clusters.

As an example, Fig. 3 is reporting the genotype-calling polar coordinate graph of 190 samples for a selected SNP. Samples are default coloured according to their called genotypes and the number of data points in each cluster is indicated below the x-axis : sample lying within the dark red region are called AA, those within the dark purple, AB and those falling in the dark blue region are called BB (theta value equal to 1). The cluster ovals are characterized by a diameter of two-standard deviations. Sample quality assessment is performed through the GenCall score (GC) which reflects the distance of a data point to a certain cluster's centroid and represent the confidence of that particular genotype calling. A threshold (typically GC ≤ 0.15 in Infinium experiments, dark shaded regions in Fig. 3) is applied to filter poor quality data which will be declared as missing (pale shaded regions correspond to the no-call regions) In addition, sample-averaged GC scores are used to check SNP data quality.

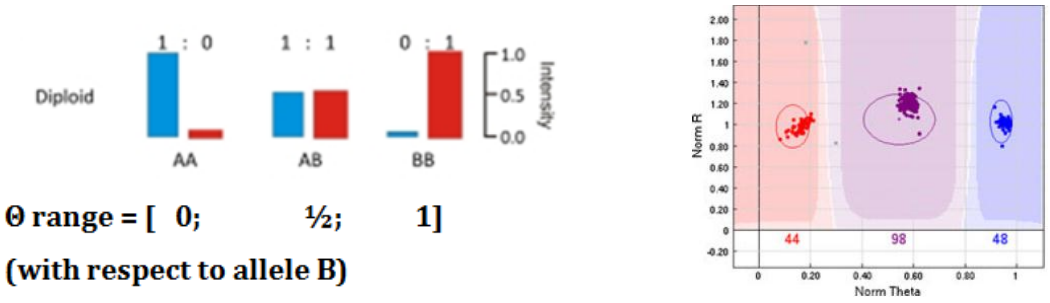


Fig. 3 On the left: the expected signal intensities for AA, AB, and BB genotypes in a diploid species (and relative θ values). On the right: polar coordinate graph regarding a SNP typed in a diploid species.

The 90k BeadChip utilized for NCCR genotyping was composed by 81,587 gene-associated SNPs in total, the majority of them originally identified after trascriptomic analysis of bread wheat elite varieties. Around 8,000 durum wheat SNPs contributed by CRA and the AGER project (PSB, UniBO, UniUD). We evaluate the transferability of bread wheat SNPs in durum wheat with a pilot study carried out using a bread wheat-derived 9K SNP Illumina Infium array to analyse 43 tetraploid wheat genotypes. Approximately 44% of the assayed SNPs resulted polymorphic (i.e. any SNP test showing at least one reliable polymorphism and <3 missing calls among the 43 tested genotypes) while the proportion of detected real informative (with MAF>10%) SNPs was lower, accounting for approximately 15-16% of the total assay. Around half of such informative set of SNPs behaved as diploid while the rest appeared as reliable hemi-SNPs (i.e. a signal is produced by both homoeolog marker loci whilst only one homoeolog couple is polymorphic). The pilot study therefore proved the adequacy of using SNP platforms developed with SNP discovery efforts in bread-wheat to genotype the less characterized tetraploid wheat. On the other hand, a consistent decrease in the level of parallel SNP typing has to be expected.

The effective genotyping of NCCR population with the 90k Illumina array was performed by the Illumina Provider Trait Genetics (M. Ganal, Gatersleben, Germany).

2.3 Cluster file development for allele calling in durum wheat

The Illumina GenCall algorithm embedded in the Genome Studio software has been optimized for allele calling in diploid organisms and it has been trained to have an *a priori* expectation for three possible genotypic states. As previously described, the optimus cluster positions for genotypes AA, AB and BB are corresponding, in polar coordinates, to theta equal to 0, 0.5 and 1. The allotetraploid genome of durum wheat might determine significant deviations from standard diploid allelic ratios hampering sample clustering (Akhunov *et al.*, 2009). With

reference to the polar genotyping graph in Genome Studio (in the normalized theta/R plot), such situation could be exemplified by a cluster compression towards the theta values of 0, 0.25 and 0.5 or , referring to the other allele: 0.5, 0.75 and 1.0 (see Fig. 4).

In addition, paralogous loci and possibly occurring null alleles might also contribute to interfere with the allelic dosage of precise SNPs and a maximum of five (instead of three) possible genotype classes are possible to be expected at each marker position (E. Akhunov *et al*., 2009). Indeed, the ploidy nature of durum wheat made not feasible automated GenCall calling for those assays producing multiple clusters, compressed clusters or low intensity signals (Cavanagh *et al*., 2013).

A dedicated cluster file was therefore produced by Trait Genetics and UNIBO with the aim of assuring correct capturing of variation in data also in the presence of genotyping plots characterized by shifted SNP clusters. Editing the new cluster file was quite time-consuming since we had to manually adjust the cluster positions and sizes according to the performance of the 81,587 SNPs on durum wheat samples. The calibration of the cluster positions to durum wheat data was performed on the Genome Studio normalized theta/R plot after applying the conventional GenCall diploid version.

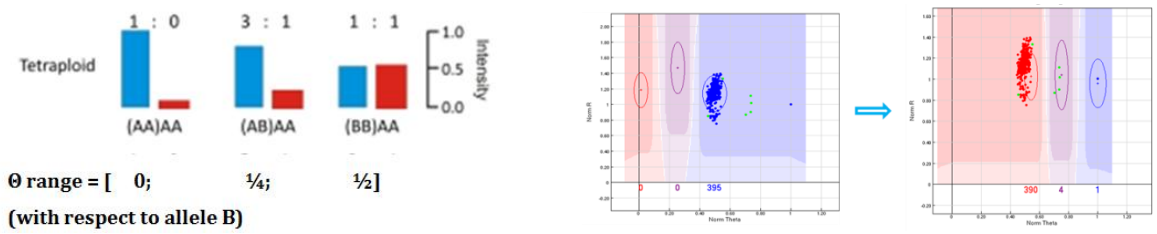
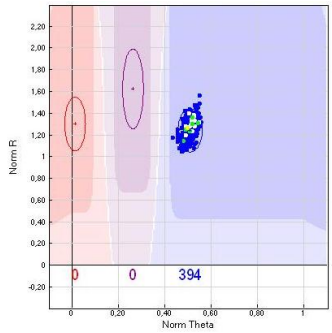


Fig. 4 On the left: the expected signal intensities for (AA)AA, (AB)AA, and (BB)AA genotypes for an hemi-SNP typed in a tetraploid species (and relative θ values). On the right: polar coordinate graph regarding a SNP typed in a diploid species.

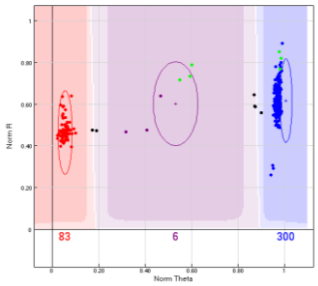
The training durum wheat materials involved: a diverse panel of durum accessions, NCCR RILs ,six F_1 samples (Dylan x Normanno; Tiziana x Normanno; Dupri x Normanno; Achille x Normanno; Strongfield x Saragolla; Kofa x Claudio) and their corresponding 9 F1 parental lines. The presence of F1 samples was particularly important to assess reproducibility of the assay.

In relation to the three major classes of SNPs which have been described in durum wheat (see Introduction), the possible scenarios which could be encountered while analysing durum wheat samples through the GenCall algorithm were:

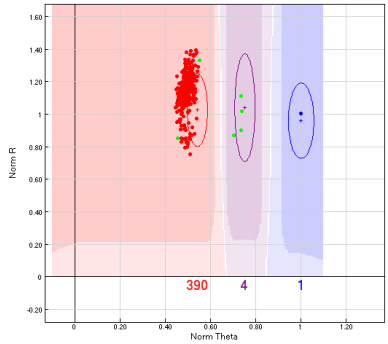
1) Inter-homeologous SNPs: non-informative SNP class, the marker is discarded.



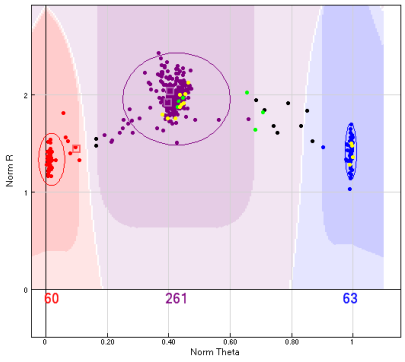
2) Simple SNPs (genome-specific SNPs): no need to edit the cluster positions as they generally behave as an ordinary SNP test in diploid organisms.



3) Hemi-SNP: e.g. with a fixed AA homeolog: (AA/AA), (AB/AA), (BB/AA), Θ (with respect to the B allele)= [0; ¼; ½]. Cluster shifting, editing needed.



4) Paralogous/duplicated loci could cause the simultaneous detection of more-than-3 clusters: Θ (with respect to allele B)= [0; ¼; ½; ¾; 1]. The marker is discarded.



The finally produced durum wheat-dedicated cluster file resulted able to score 25,631 polymorphic markers (5,946 genome-specific and 19,685 unspecific assays). The monomorphic markers were 50,468 while the overall failed assays were 5,488.

3. Map construction

After eliminating markers characterized by > 10% missing data we examined NCCR for potential segregation distortion. In NCCR two types of segregation patterns were present due to the founder alleles distributions, namely 1:1 (e.g. 0-1-0-1 or 0-0-1-1) and 1 : 3 (e.g. 0-1-1-1 or 0-0-0-1). Markers were therefore filtered based on the expected RIL values, with an allowed distortion of +/- 30% (corresponding to a skewed segregation p-value < 10^{-4}). A more stringent threshold (+/- 10% allowed distortion) was adopted post-reconstruction of parental information.

The NCCR linkage map was estimated using the program mpMap (Huang and George, 2011), an R package specifically written for the analysis of MPPs. Maximum likelihood estimates of recombination fraction (r) between each SNP pair were obtained using the *mpestrf* function with default parameter settings. We then clustered the markers in linkage groups when the pairwise LOD score exceeded 5.0, through the *mpgroup* function set to form 50 initial groups. We explored the LOD and recombination fraction matrices with interactive heatmaps in R (R Core Team 2013) and iteratively collapsed the number of marker clusters to 22 based on closest linkage between clusters. The 22 linkage groups were then combined into 14 groups corresponding to the *T. durum* chromosomes by leveraging on markers in common with a

recently developed *T. aestivum* consensus 90k SNP map (Wang *et al.*, in press), combining six different bi-parental maps and comprising 40,266 SNP markers typed with the same 90k SNP Illumina array used in this population. Markers were ordered within each linkage group using the *mporder* function which to minimize the sum of adjacent recombination fractions. We estimated map positions using the function *computemap* (Haldane map function, maxOffset = 1) imputing missing r values based on values at the most highly correlated marker.

The final map has been described and validated through the following methods.

Method IV of Chakravarti *et al*. (1991) was used to adjust the length of each linkage group and thus the total length of the genetic map. The marker coverage (c) of the map was estimated by the formula $c = 1\text{-}e^{-2dm/L}$ (with 2d = distance between adjacent markers, L = cM length of genome; m = number of mapped markers. (Lange and Boehnke 1982; Remington *et al.*, 1999) under the assumption of random distribution of markers over the LGs.

The genome coverage was further investigated by taking as a reference the previously cited *T. aestivum* consensus 90k SNP map (Wang *et al*., 2014, in press) and the local deviations from the marker random distribution that frequently occurred along the LG. Based on the projection of the NCCR mapped markers on the hexaploid wheat consensus 90k SNP map (considering reference map intervals of 5 cM), the observed A and B genome coverage of the NCCR map was calculated for each chromosome. Additionally, Spearman's rank correlation values between the two considered maps have been calculated for all chromosome pair comparisons.

4. Haplotype reconstruction

Founder probabilities were computed with the *mpprob* function in mpMap using an Hidden Markov model implemented in R/qtl (Broman *et al.*, 2003). Haplotype were then reconstructed by imputing founder alleles as known if the corresponding probabilities at a genomic location exceeded 0.7. The original bi-allelic SNP data were thus recoded on the basis of the most likely founder haplotype at each particular linkage block, increasing the overall marker information content. Recombination events were declared at loci where the founder allele changed along the genome. In addition, markers co-located along the genome were removed in order to minimize redundancy. The computation of founder probabilities for each RIL along the chromosomes enabled us to estimate founder-specific additive genetic effects at each QTL.

5. Field experimental design

The phenotypic evaluation of NCCR population was carried out during two harvesting years (2010-2011 and 2011-2012) in three similar locations in the Po Valley: Cadriano (44°33'N lat., 11°24'E long. 2010-2011 and 2011-2012. Sizes of plots were, respectively, 2.4 m^2 and 2.28 m^2); Poggio Renatico (44°45'57''N, 11°25'31''E, 2010-2011. Plots: 4 m^2) and Argelato (44°39'03'' N, 11°20'34'' E, 2011-2012. Plots: 4 m^2). In each environment, as control genotypes, we inserted three replicates of the cultivars Levante, Meridiano, Orobel, Saragolla and Svevo, as well as the four parental lines replicated twice. The 338 RI lines, the four parents and the five control genotypes were evaluated according to an incomplete-block experimental design. A 19 x 19 α-lattice design (Petterson and Wiliams, 1976) with two replications was considered for each environment blocking and was designed by means of FACTEX procedure (SAS Institute, 2006).

6. Phenotypic evaluation

NCCR population was phenotypically evaluated in open-field except for grain yield and yield-related traits and seed quality traits (Grain protein content) which were analysed in the post-harvest phases. Where not specified, traits were recorded in all environments(four). The phenotypic traits were:

- **Phenology-related traits**

1) Heading date (HD, days to heading from sowing) was recorded when 50% of the ears in a plot showed emergence out of the flag leaf.

2) Maturity date (MD, days to maturity from sowing) was determined when 50 % of the peduncles in a plot have totally turned to yellow color. Data for MD in Poggio Renatico was not available.

- **Morpho-pysiological traits**

3) Early ground cover (EGC, visual score). EGC it is an early-vigour measure which indicates the strength of plants in early stages, an important feature above all in low input environments.

4) Flag leaf erectness (FLER, visual score estimated at booting stage, Zadok 47).

5) Plant height (PH, cm). PH was measured at maturity from ground level to the top of the terminal spikelet (excluding awns) on four main culms per plot. Data for PH in Argelato was not available.

6) Flag leaf greenness (FLG, SPAD unit). The measured SPAD index was the average of three readings effectuated at beginning of anthesis (Zadok 60) on 20 representative flag leaves by means of a SPAD-502 chlorophyll meter (Konika Minolta). The SPAD index measures the canopy spectral reflectance and is generally used to assess the chlorophyll content of leaves *in situ.* Until flowering, SPAD readings are generally correlated to yield potential while, during grain filling, represents a "greenness" measurements which can estimate plant senescence. As much of the N in a leaf is partitioned in chlorophyll, SPAD index could be also used to monitor the nitrogen status of crops (Filella *et al*., 1995).

7) Normalized difference vegetation index (NDVI, NDVI units). The NDVI is a canopy spectro-radiometric index capable of assessing green biomass and plant nitrogen (N) content. The canopy spectral reflectance is estimated as: NDVI =(NIR-R)/(NIR+R), where NIR=value of near-infrared radiation from a pixel and R= red wavelengths of the whole range of PAR from a pixel (Value of Phtosynthetically Active Radiation). NDVI was measured at three different Zadok's scales: Zadok 31 (first node detectable, stem elongation stage), Zadok 40 (booting stage) and Zadok 57 (emergence of the ear from the boot).

- **Yield and yield-related traits**

8) Fertile Tillers * m^{-2} (FTsm, number). Number of fertile tillers (at least one grain on the ear) considering one square meter for each plot.

9) Grain volume weight (GVW, $Kg*hl^{-1}$). Measured by a Grain Analysis Computer, GAC 2100 (Dickey-John Corporation, Minneapolis, MN, USA). On the same sample of grains, the kernel moisture (%) have been recorded.

1) Thousand kernel weight (GWT, g/1000 kernels) was recorded averaging the test weight values of two samples of 50 kernels for each line.

2) Spikelet number per ear (SNE, number). Potential spikelet number was measured on six selected ears for each line. Trait recorded in Cadriano 2010-2011 and 2011-2012 only.

3) Grain number per ear (GNE, number). Grain yield per spike was measured on six selected ears for each line. Trait recorded in Cadriano 2010-2011 and 2011-2012 only.

4) Grain number per spikelet (GNS, number). Derived from the ratio GNE/SNE per ear. Trait recorded in Cadriano 2010-2011 and 2011-2012 only.

5) Grain weight per ear (GWE, g). Average grain weight of six selected ear for each line. Trait recorded only in Cadriano, 2010-2011 and 2011-2012.

6) Grain yield (GY, t ha-1). Grain production per plot was adjusted accounting for sample moisture percentages and converted to tonnes per hectares.

Quality traits

7) Grain protein content (GPC, %). The percentage of protein content on total dry weight was obtained with near-infrared reflectance spectroscopy.

7. Statistical analysis of phenotypic data

Analysis of variance per environment was performed with the LATTICE procedure (SAS Institute Inc., 2006), according to Cochran and Cox (1960). Genotype's least square means were adjusted for lattice if its relative efficiency was greater than 105% when compared to that of a randomized complete block design. The ANOVA over environments was performed on the basis of the least square means of each environment (i.e. location x year) *via* the SAS GLM procedure. After weighting the ANOVA for replications in each location, the residual pooled over environments was considered as the error term. Heritability for each trait was estimated on the adjusted means calculated across the four environments (n) and the two replications (r) per environment, as: $h^2 = \sigma^2_{g/(} \sigma^2_g + \sigma^2_{ge}/n + \sigma^2/nr)$ and on the single environment data as: $h^2 = \sigma^2_g$ /

$\sigma^2_g + (\sigma^2_{error}/r)$. The descriptive statistics and the correlation values (Pearson's correlation coefficients) were calculated in R (R Core Team 2013).

8. QTL mapping

Composite interval mapping was conducted to detect QTL using mpMap, an R-coded platform specific for analysing multiparental populations (Huang and George, 2011). Accordingly, marker-trait statistical associations are evaluated accommodating the conditional founder haplotype probabilities at each locus, given the available marker genotype data. Therefore, mpMap CIM QTL analysis will be further denoted as IBD-CIM.

The phenotypic dataset was composed by combined adjusted means over years, environments replications nested in environments. For a subset of phenotypic data (HD, MD, PH, GY), the QTL analysis was carried out by two different strategies, namely single marker analysis with Tassel Version 3.0 (Bradbury *et al*., 2007) and composite interval mapping with mpMap (Huang and George, 2011). Moreover, for the same subset of data, we have carried out also independent QTL analysis for each of the four location*year combinations, namely Cadriano 2010-2011 (Cad11), Cadriano 2011-2012 (Cad12), Poggio Renatico (Pr11) and Argelato (Arg12). These single-environment QTL analysis were run with mpMap on the adjusted means calculated across the two replications per environment.

With mpMap, we performed interval mapping using the mpIM function (program 'qtl'), fitting in ASReml v3.0 (Butler et al. 2007) a linear model and separately estimating fixed effects for each of the four founders along the chromosomes (Rascon/Tarro has been arbitrarily chosen as the reference haplotype). The regression was performed at each position for which we estimated founder probabilities. We first ran simple interval mapping (SIM) and consequently fit the number of SIM-detected QTL in a composite interval mapping (CIM) (Jansen and Stam, 1994; Zeng, 1994) model which accounted for background variation. CIM was estimated on the reduced map representing only unique genomic positions. A Wald statistic testing the overall significance of all founder effects was computed at each hypothesized QTL position. In addition, to estimate the amount of putative functional alleles at each identified QTL, t-test were performed on the founder effects. After Bonferroni's correction, a *p-value* threshold of 0.05 was applied to each founder's significance. We considered a *p*-value equal to 10^{-3} as marker-wise detection threshold for putative QTL. The reported founder effects and the full model R^2 were obtained by fitting all QTL simultaneously. The percentage of phenotypic variation accounted for

by each individual QTL (R^2) was determined as the square of the partial correlation coefficient, fitting the final multiple regression model. We finally calculated the LOD-2 supporting interval (for combined data) and the LOD-1 supporting interval (for single-environment data) based on transformed *p-value* [−log10(*p*)] profiles. QTL clusters have been determined at chromosome regions where QTL for different traits showed overlapping supporting and/or significance intervals.

In Tassel, we utilized a General Linear Model (GLM) analysis and performed an *F*-test at each SNP position to conduct a single marker analysis (SMA). When testing for the presence of a putative QTL, SMA test was merely considering the bi-allelic information present at a certain position. As a result, the QTL analysis performed with Tassel will be further indicated as IBS-SMA. Putative QTL have been identified applying a marker-wise threshold of 10^{-3} to the results of the genome-wide associations obtained with Tassel. In addition, only QTL supported at least by three markers significantly associated with the trait have been considered as true. For each QTL, the peak was identified in correspondence with the marker showing the highest LOD value in the target interval. We then computed QTL significance intervals by a sliding window applied to SMA LOD values (± 4 marker, threshold at LOD equal to 1.44).

The putative QTL detected by SMA have been then additionally checked for robustness comparing their peak *p*-values with the marker-wise p-value corresponding to a genome-wise significance level of $\alpha = 0.05$. We therefore calculated the effective number of independent tests, $M_{eff,}$ in order to correct such genome wise threshold for multiple testing (Cheverud, 2001; Nyholt, 2004). This procedure has been previously shown to significantly smooth conventional Bonferroni adjustment which is often overly conservative (Gao *et al*., 2010). The number of independent tests was inferred by Haploview, defining the number of LD blocks with the four-Gamete rule and setting the cut off for examining haplotypes at 20%. The estimated number of independent statistical tests had a mean of 30 over the 14 durum chromosomes and has been approximated genome-wide by 500 linkage blocks. We then obtained 10^{-4} as marker-wise significance threshold for claiming a statistical SNP-phenotype association corresponding to the target genome-wise significance level of $\alpha = 0.05$.

RESULTS

1. Genetic diversity represented by NCCR founders

The 90k SNP genotyping array was utilized to genotype a collection of 280 durum wheat cultivars and accessions available at Unibo (Maccaferri *et al*. in preparation), which included the four NCCR founders. The large dataset were preliminarily analysed in order to evaluate the genetic diversity explored by the NCCR cross in the context of cultivated durum wheat. According to the UPGMA analysis (Fig 1), the four NCCR founders seem to represent quite well the major breeding groups of elite durum wheat, as also delineated by Maccaferri *et al*. (2003). More precisely, while Claudio and Colosseo both belong and were classified into the Italian group of accessions, they actually represent different genetic sub-pools, as Claudio is a recent Cimmyt-derived cultivar. The only important gene pool clearly not represented by the four NCCR founders is the Icarda one (which includes the subpools 'Icarda temperate areas_Cham1' and 'Icarda drylands_Omrabi3'. See Fig. 1)

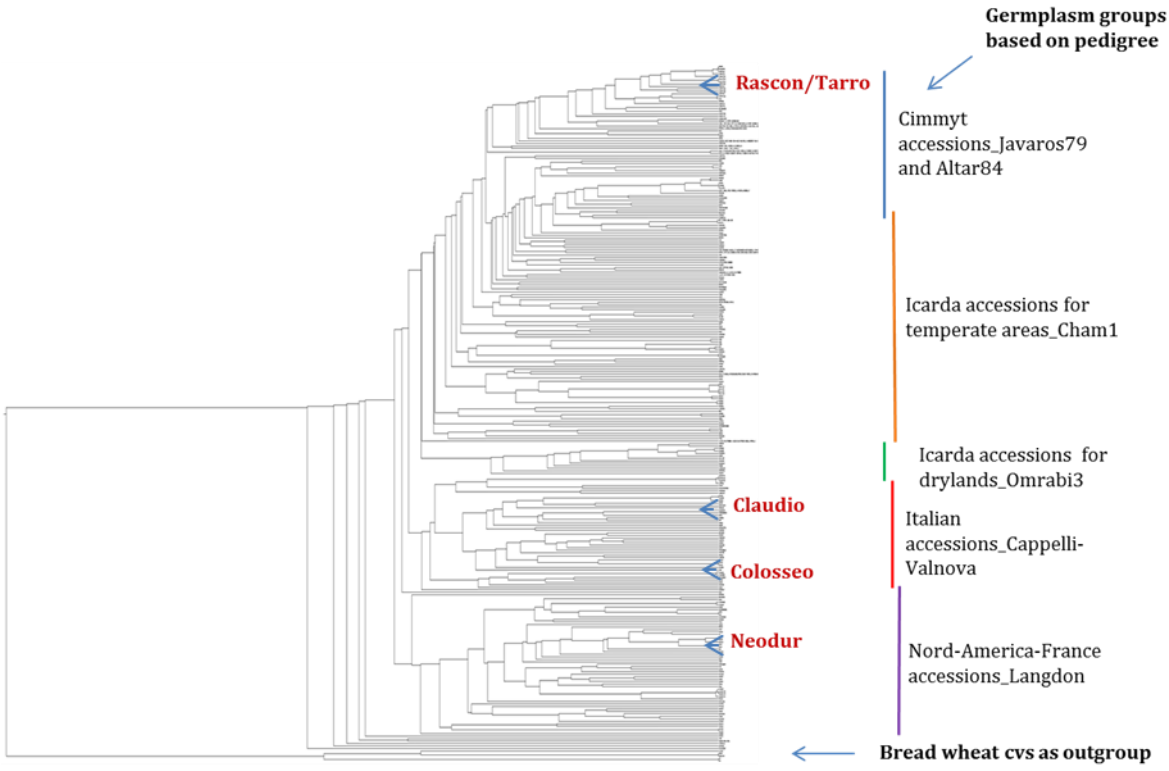


Fig. 1 Dendogram representing the UPGMA cluster analysis of 280 durum wheat cultivars and accessions (Tassel v.4) based on SNP data produced by the wheat-dedicated 90k Illumina chip.

2. 90k SNP chip performance on NCCR

The application of the 90k ILLUMINA SNP array and the specifically developed cluster file (for correct genotyping of tetraploid samples) to NCCR enabled us to genotype the 338 NCCR RILs and the four founders by means of the 81,587 SNPs present in the chip.

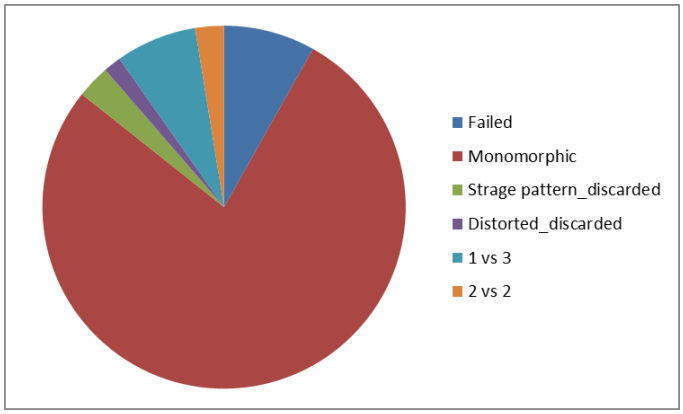


Fig. 2 Performance of the 90k SNP array on the NCCR population. Failed = SNP with >10% of RILs with missing genotype calls; Monomorphic = monomorphic data; Strange pattern_discarded = ambiguity in the founders' data; Distorted_discarded = segregation distortion with reference to the expected RIL values; 1 vs 3 = proportion of SNP showing 1 to 3 allele segregation; 2 vs 2 = proportion of SNP showing 2 to 2 allele segregation.

Approximately 78% of such marker set resulted monomorphic in NCCR (Fig. 2). After filtering the remaining dataset for missing information and distortion in marker segregation pattern, we obtained the final polymorphic SNP set (7,959 markers) available for developing the NCCR linkage map. The percentage of markers showing a segregation pattern consistent with the model 1:1 was 0.26% while the amount of markers segregating as 1:3 was equal to 0.7%.

3. NCCR genetic map

The final number of SNPs included in NCCR linkage map was equal to 7,594. The map covered the 14 durum wheat chromosomes and spanned 2,663 cM, distributed as 1,188 for the A genome and 1,475 for the B genome (details reported in Fig. 3 and Table 1). Collapsing the markers to unique positions resulted in 1,229 co-segregating clusters of markers. The number of unique positions for each chromosome varied from 43 (chr 1A) to 134 (chr 1B) and had a mean of 87.8.

The total estimated genome length of the NCCR map (after adjustment following Chakravarti *et al.*, 1991) was 2,674.25 cM. The overall average marker distance was 2.8 cM. The histogram of inter-marker distances (Fig. 4) showed the large majority of pairs were in the class of 0-1 cM (6,365 over 7,580, 84.0%). However there were 135 inter-marker distances of 10-50 cM, indicating potential gaps in optimal marker coverage.

We further investigated the genome coverage by taking as a reference the recently published 90k SNP consensus map for *T. aestivum* (bread wheat) (90k bread. Wang *et al.*, in press). A monotonic increase was detected in NCCR *vs* 90k bread scatterplots (Fig. 6) with an extremely high conservation of marker order (Spearman's rank correlation >0.95 for all chromosome pair comparisons). Based on the same comparison, the A and B genome coverage of the NCCR map was calculated for each chromosome and is reported in Fig. 5. Gaps in genome coverage as referred to the hexaploid wheat consensus map were comprised between 5 and 14% for most of the chromosomes and the majority of the gaps were accounted for by the distal telomeric regions. The observed interstitial gaps did not extended over 5% of the consensus map. Total gaps equal or lower than 2% were observed for four chromosomes (2A, 6A, 6B, 7B). On the contrary, chromosomes 4A and 5A showed large non covered regions of 22.4 and 42.7%, both in the distal chromosome regions.

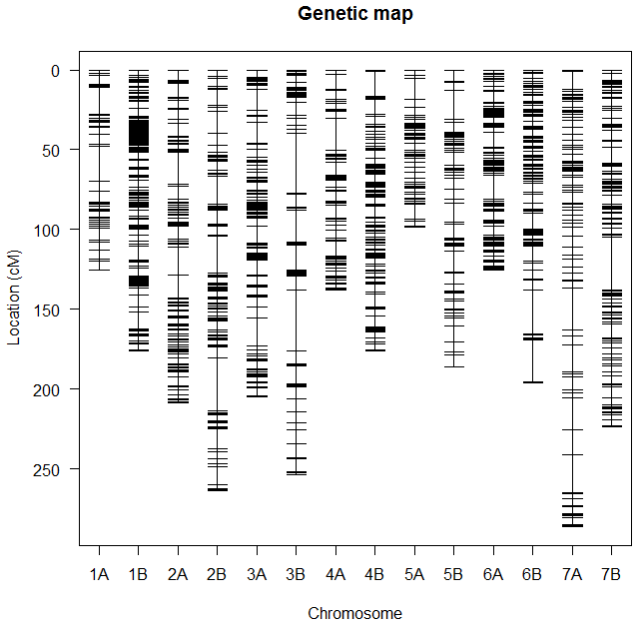


Fig. 3 The NCCR linkage map.

Table 1 Summary of main parameters about the NCCR map by chromosomes. For each chromosome is reported: number of SNP markers, length (cM) of the chromosome, number of unique positions (bins of co-segregation among markers) and average number of recombination events for each RI line.

Chr	No. SNPs	Lenght (cM)	Unique Pos	No.Rec/chr
1A	281	125.2	43	1.8
1B	864	176.2	134	2.4
2A	695	208.8	99	3.0
2B	542	263.7	101	2.4
3A	622	204.8	104	3.3
3B	304	253.7	64	1.2
4A	359	138	68	1.7
4B	552	175.9	105	2.6
5A	177	98.8	53	1.6
5B	408	186.4	65	2.3
6A	708	125.7	91	1.5
6B	719	196	101	2.7
7A	632	286.4	82	3.5
7B	731	223.6	119	1.6
Tot	7,594	2,663.20	1,229	31.5

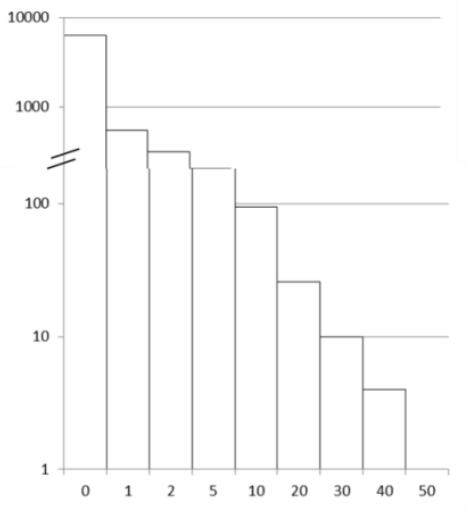


Fig. **4** Classes of inter-marker distances(in cM).

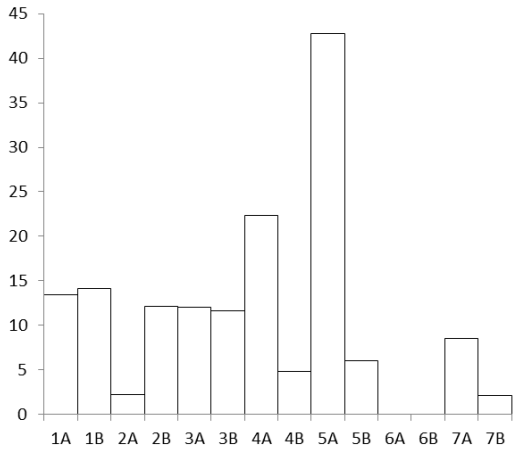


Fig. 5 Chromosome coverage of the NCCR map, calculated projecting NCCR mapped markers on the bread wheat reference map, considering intervals of 5 cM.

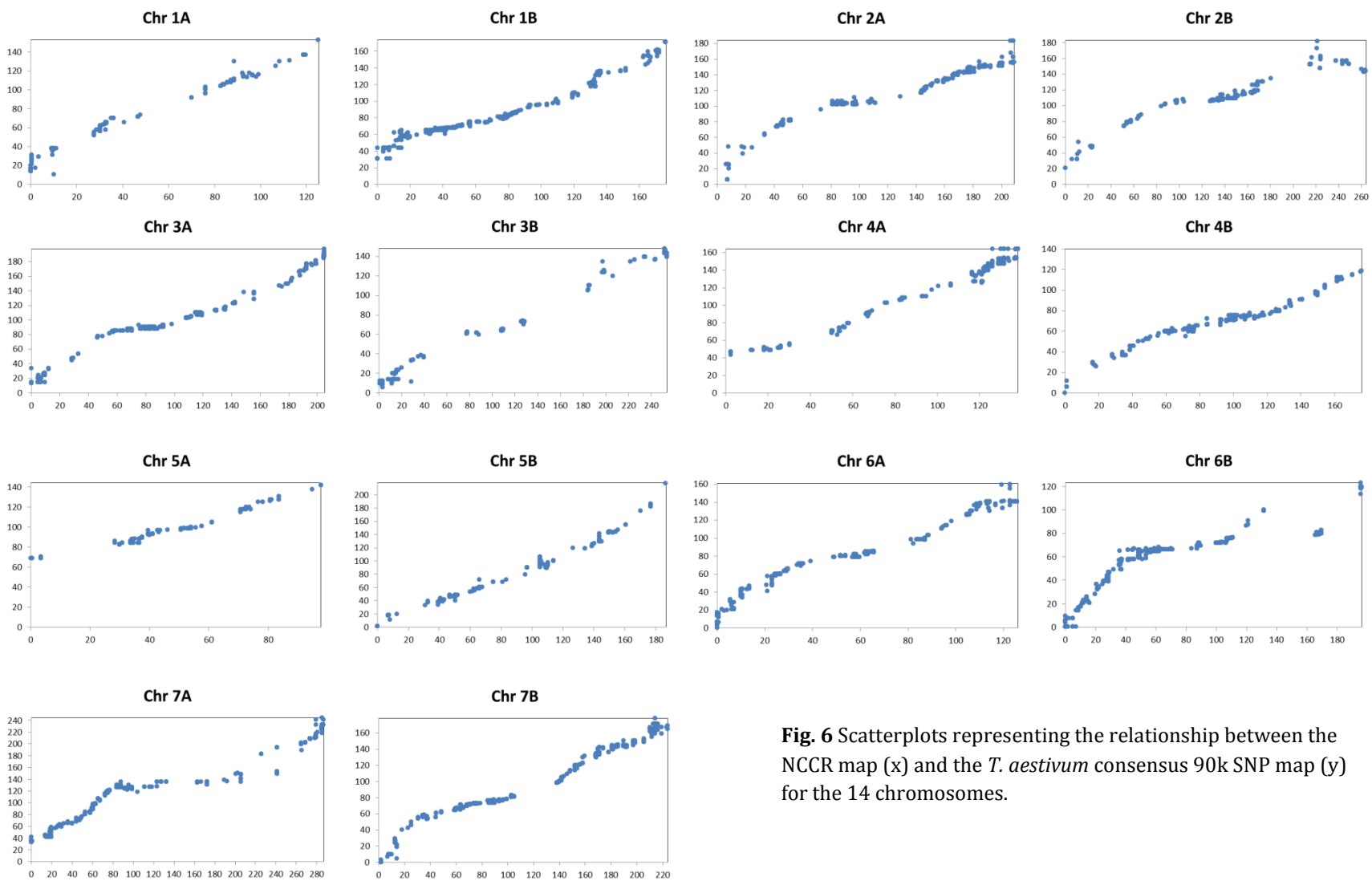


Fig. 6 Scatterplots representing the relationship between the NCCR map (x) and the *T. aestivum* consensus 90k SNP map (y) for the 14 chromosomes.

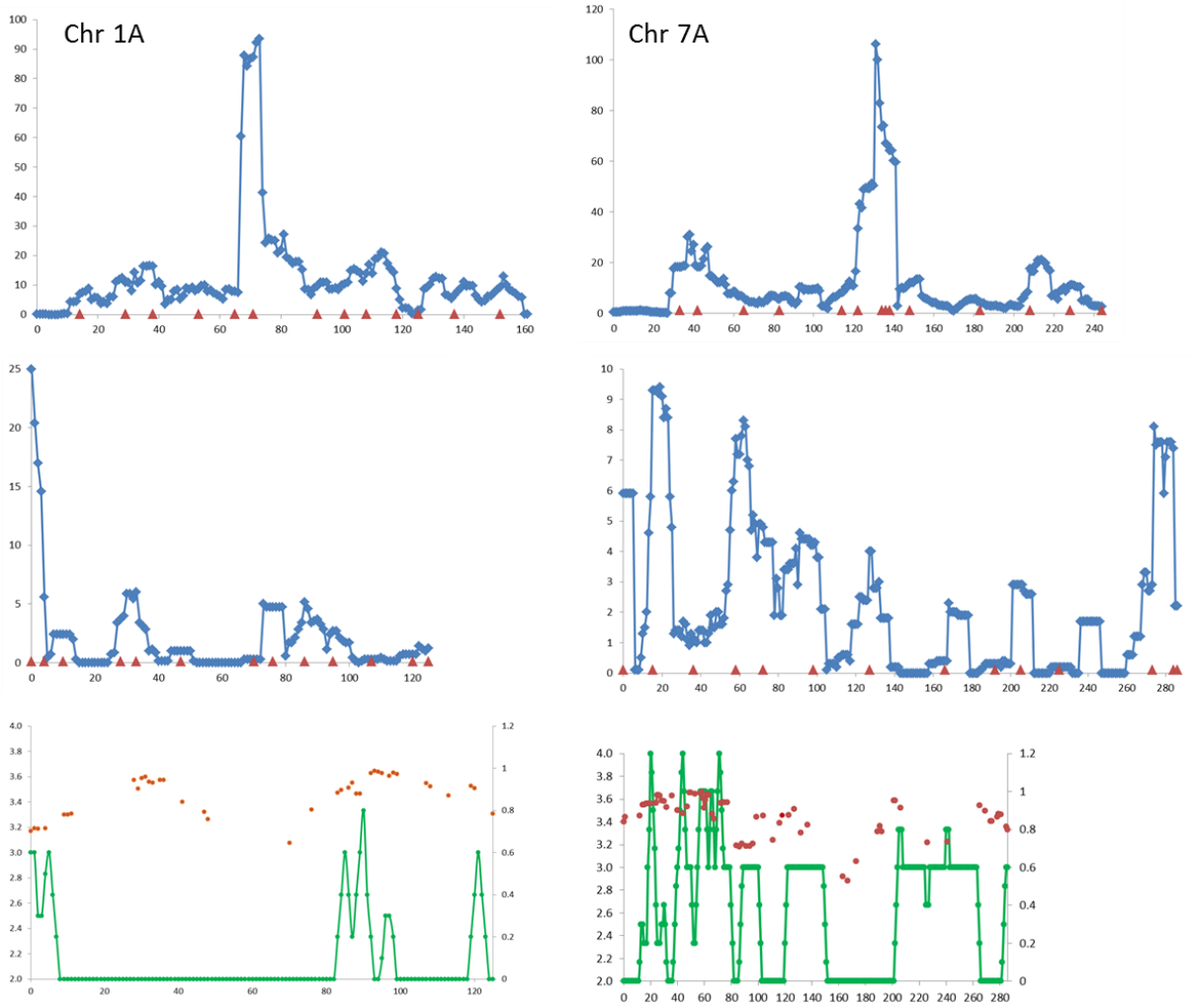


Fig. 7 From top to bottom: 1) SNP coverage of the consensus 90k map, 2) SNP coverage of NCCR map. Red triangles represent anchor markers. 3) NCCR map: 3cM-sliding window of the haplotype density along the chromosome (in green) and probability of assignment to a certain founder at the 0.7 target threshold (red dotted scatterplot). Study cases of chromosome 1A (on the left) and chromosome 7A (on the right).

4. Genetic structure of NCCR population

Estimating the founder probabilities for each RIL along the genome allowed a better understanding of the genetic structure of the population. Average assignment percentages of each founder along the 14 chromosomes ranged between 5.5% (Claudio haplotype, chr 3B) and 30.4% (Neodur haplotype, chr 5A) and had a genomic mean of 20% across the four founders (compared to 25% expected). The percentage of regions where founders could not be assigned per chromosome varied from 9.0% (chr 5A) to 46.1% (chr 3B). On average, for each line, 20.1% of the genome could not be assigned at the chosen threshold probability (0.7) to any founder. Average genome-wide founder assignment percentages for all the 14 chromosomes are reported in Fig. 8. Percentage of founder assignment was typically lower at centromeric regions, which also showed a lower SNP density and lower haplotype diversity between the four founders (Fig 8). The genome-wide average number of recombination events per line was 31.5, while the average number of recombination events per chromosome varied from 1.2 (3B) to 3.4 (7A), see Table 1.

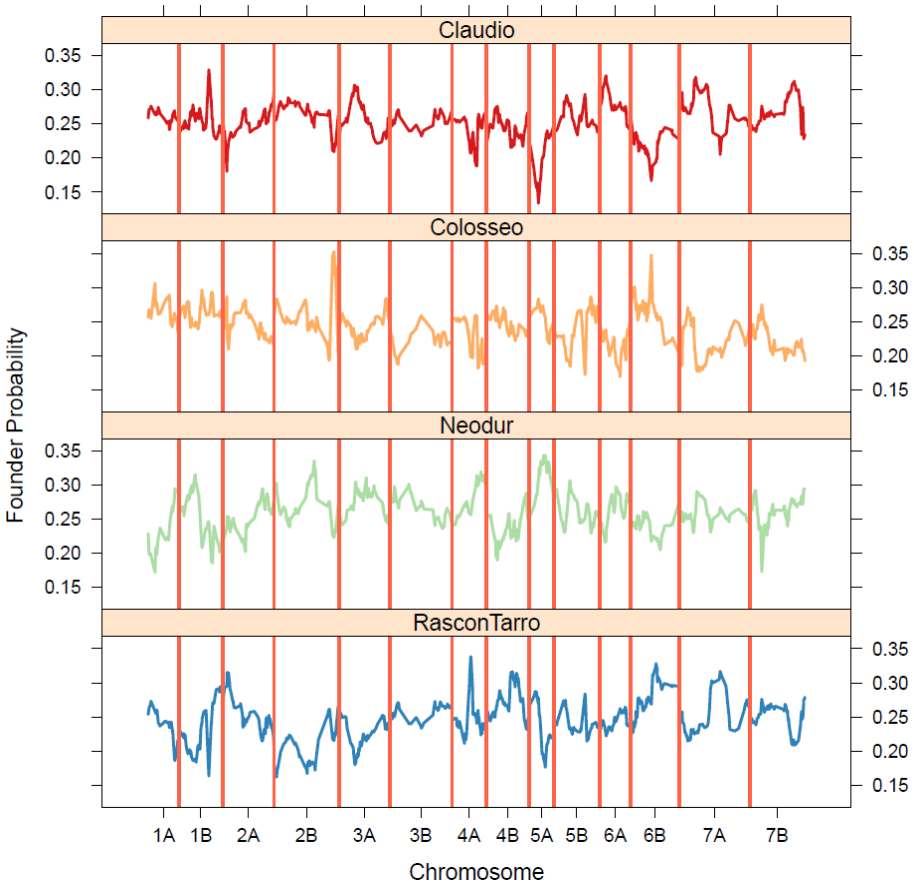


Fig. 8 Percentages of genome-wide founder assignment for all chromosomes. Orange lines indicate boundaries of chromosomes.

5. Statistical analysis of phenotypic data

In the NCCR population, all traits followed a normal distribution (excluding a skewed distribution identified in the FLER-score dataset) and were characterized by transgressive segregation with respect to the parental lines (Fig. 9). In particular, a clear positive transgressive segregation has been observed for EGC, NDVI_31, and GPC. Colosseo was the earliest (both in terms of days to heading and maturity) and lowest yielding parent, Neodur was the latest and Claudio was the highest yielding (Fig. 9). The four founders appeared characterized by favourable alleles at different yield-potential QTL, consistently with the expectations (see Materials and Methods). With respect to the parental values, Neodur showed the highest SNE values, Claudio the greatest GVW and Ftsm values, Colosseo the highest GWT while Rascon/Tarro the highest GNS.

The general linear model highlighted highly significant differences among NCCR RIL genotypes and highly significant genotype-environment interactions for the target traits (p-values $<<10^{-4}$, see Table 4). Broad sense heritability for the combined values over the four environments returned values higher than 90% for HD (94.5%) and GWT (93.6%). GY heritability was estimated as 47.7% (Table 4). GVW, for which the h^2 value was estimated as equal to 25.4% in the present experiment, was discarded from further analysis.

The Pearson's correlation values regarding the 18 traits herein considered are reported in Table 2 (combined phenotypic data) and Table 3 (single-environment data for a subset of traits: HD, MD, PH, GY). With combined data a positive and relatively high correlation was found between HD and MD (r = 0.62). A positive HD-MD correlation have been encountered as well over single-environment data (Cad11=0.63; Cad12=0.33; Arg12=0.45). Regarding NDVI measurements, N31 showed an high correlation with N40 (0.77) and a slightly lower correlation value with N57 (0.60). The correlation between N40 and N57 was estimated as r = 0.74. The three NDVI measurements permitted to identify different correlation values with regards to the other examined traits. As expected, N 31 showed a considerable correlation (r = 0.68) with EGC, while the magnitude of such correlation seems to decrease with the progressing of plant development (N40 and EGC, r = 0.44; N57 and EGC, r = 0.32). An opposite whilst milder trend has been identified for NDVI31, 40 and 57 and PH (N31 and PH, r=0.33; N40 and PH, r = 0.39; N57 and PH, r = 0.46). Additionally, N40 was weakly correlated with FLER (r = 0.42) and GY (r = 0.31), while N57 showed a sizable correlation with HD (r = 0.53) and GY (r = 0.40). Combined PH data appeared to have a weak and positive correlation with GWE (r = 0.35) and GY (r = 0.33). On single-environment data, PH significantly correlates only with HD in Cad11. Considering yield

numerical components, HD appeared to positively correlate with SNE (r = 0.46), GNE (r = 0.38) and GWE (r = 0.40), MD with SNE (r = 0.38). SNE showed a positive correlation with GNE (r = 0.38) and GWE (r = 0.45). Being a GNE-derived measure, GNS was obviously extremely correlated with GNE (r = 0.90). GNS then presented a moderate positive correlation with GWE (r = 0.52) and a negative correlation with GPC (r = -0.54). GNE resulted as highly and positively correlated with GWE (r = 0.68) and negatively with GPC (r = 0.59). GY appeared as slightly correlated with GNE (r=0.35), GNS (r = 0.32) and GWE (0.42). GWT appeared to have a relatively strong negative correlation with the numerical yield components GNE (r = -0.53) and GNS (r = -0.59) and an r value of 0.39 with regards to GPC. Considering the spike-yield value GWE, a negative and significant correlation (r = 0.32) was detected with regards to GPC .

Tab. 2 Correlation (Pearson's) values among phenotypic traits collected on NCCR. Significant correlation values > |0.30| are shown bolded excel boxes.

	HD	MD	EGC	FLER	PH	FLG	N31	N40	N57	FTsm	GVW	GWT	SNE	GNE	GNS	GWE	GY	GPC
HD	**1**																	
MD	0.62	**1**																
EGC	-0.01	0.05	**1**															
FLER	-0.06	-0.1	0.06	**1**														
PH	0.16	0.11	0.23	0.13	**1**													
FLG	-0.26	-0.13	-0.01	-0.17	0.02	**1**												
N31	0.19	0.05	0.68	0.29	0.33	-0.07	**1**											
N40	0.13	0.01	0.44	0.42	0.39	-0.08	0.77	**1**										
N57	0.53	0.26	0.32	0.27	0.46	-0.16	0.6	0.74	**1**									
FTsm	-0.14	-0.11	0.22	0.02	-0.02	-0.01	0.21	0.12	0.09	**1**								
GVW	-0.16	-0.02	0.09	-0.06	0.07	-0.07	-0.07	-0.06	-0.07	0.15	**1**							
GWT	-0.09	0.12	0.25	0.19	0.21	-0.03	0.14	0.17	0.04	-0.14	0.1	**1**						
SNE	0.46	0.38	-0.12	-0.15	0.19	-0.09	-0.15	-0.05	0.15	-0.22	0.09	0	**1**					
GNE	0.38	0.1	-0.25	-0.14	0.15	0.06	-0.15	-0.11	0.13	-0.16	-0.16	-0.53	0.38	**1**				
GNS	0.2	-0.07	-0.22	-0.08	0.07	0.11	-0.09	-0.09	0.07	-0.07	-0.22	-0.59	-0.05	0.9	**1**			
GWE	0.4	0.28	-0.09	0.01	0.35	0.05	-0.07	0.01	0.19	-0.27	-0.08	0.18	0.45	0.68	0.52	**1**		
GY	0.26	0.12	0.19	0.01	0.33	0.05	0.3	0.31	0.4	0.12	-0.02	0	0.13	0.35	0.32	0.42	**1**	
GPC	-0.21	0.04	0.2	0.07	-0.05	0.01	0.08	0.03	-0.16	-0.04	-0.04	0.39	-0.21	-0.59	-0.54	-0.32	-0.49	**1**

Tab. 3 Correlation values within single-environment data for HD, MD, PH, GY.

	Cad11						**Pr11**		
	HD_Cad11	MD_Cad11	PH_Cad11	GY_Cad11			HD_Pr11	PH_Pr11	GY_Pr11
HD_Cad11	1					HD_Pr11	1		
MD_Cad11	0.63	1				PH_Pr11	0.07	1	
PH_Cad11	0.38	0.25	1			GY_Pr11	0.34	0.14	1
GY_Cad11	0.43	0.31	0.44	1					
	Cad12						**Arg12**		
	HD_Cad12	MD_Cad12	PH_Cad12	GY_Cad12			HD_Arg12	MD_Arg12	GY_Arg12
HD_Cad12	1					HD_Arg12	1		
MD_Cad12	0.33	1				MD_Arg12	0.45	1	
PH_Cad12	0.04	0.12	1			GY_Arg12	0.26	0.26	1
GY_Cad12	-0.22	-0.15	0.45	1					

Table 4 Descriptive statistics of the phenotypic traits collected on NCCR and analysed in this work.

Trait	Abbr.	Min.	Mean	Max.	*p*- val. Tratt	h^2
Heading date	HD	132.3	140.6	149.0	***	94.4
Maturity date	MD	180.4	188.2	193.3	***	58.4
Early ground cover	EGC	3.6	6.1	7.6	***	47.4
Flag leaf erectness	FLER	1.0	2.1	5.3	***	58.8
Plant height	PH	64.0	82.1	94.8	***	86.6
Flag leaf greeness	FLG	42.6	52.8	59.2	***	63.3
NDVI_Zadoks 31	NDVI_31	0.3	0.4	0.5	***	62.0
NDVI_Zadoks 40	NDVI_40	0.5	0.7	0.7	***	51.8
NDVI_Zadoks 57	NDVI_57	0.6	0.7	0.8	***	65.6
Fertile tillers*m^{-2}	FTsm	345.6	461.2	608.8	***	32.8
Grain volume weight	GVW	73.0	77.7	81.5	***	25.4
Thousand kernel weight	GWT	38.8	51.9	63.1	***	93.6
Spikelet number per ear	SNE	14.7	17.9	23.3	***	79.6
Grain number per ear	GNE	29.3	47.7	74.4	***	78.2
Grain number per spikelet	GNS	1.6	2.7	3.9	***	74.2
Grain weight per ear	GWE	1.5	2.4	3.7	***	64.8
Grain yield	GY	4.2	5.9	7.1	***	47.7
Grain protein content	GPC	12.4	15.1	17.5	***	77.7

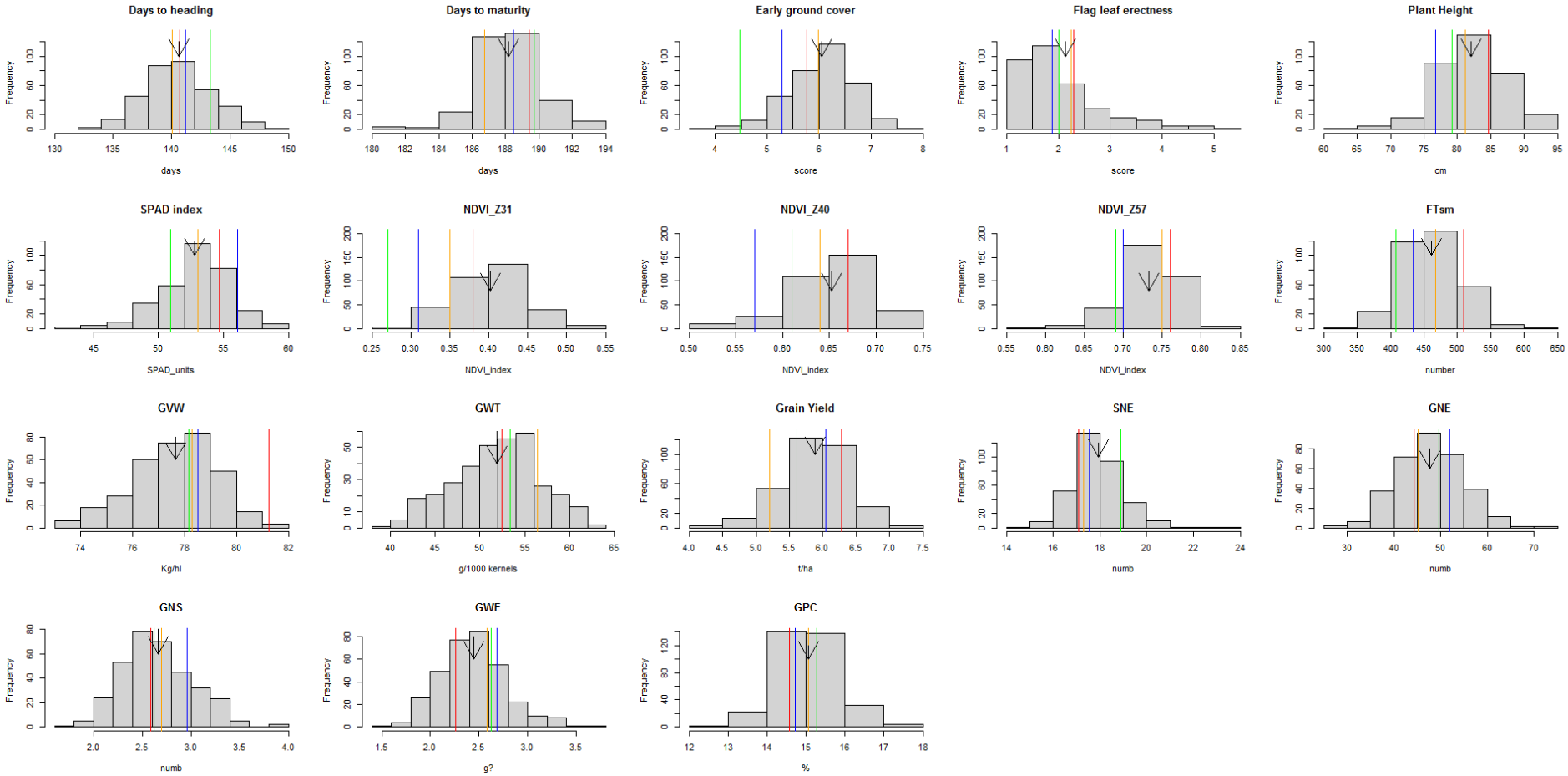


Fig. 9 Frequency distributions of the 18 phenotypic traits analysed in this work. The arrow represents the average RIL population value for the considered trait, while the parental values are depicted by colored bars (Neodur=green, Claudio=red, Colosseo=orange and Rascon/Tarro=blue).

6. QTL results

The whole-genome genetic map constructed from the SNP data was used to conduct a composite interval mapping on phenotypic data combined across years and environments.
For a subset of phenotypic traits (HD, MD, PH and GY), a comparative QTL search was conducted in parallel with two distinct methods: i) by composite interval mapping as implemented in mpMap software (hereafter reported as IBD-CIM) which exploits the four founders identity by descent haplotype probabilities, and ii) by single marker analysis using the SNP bi-allelic classes (hereafter reported as identity by state single marker analysis, IBS-SMA).

6.1 QTL mapping across environments

Considering the phenotypic data combined across years and environments, IBD-CIM identified 10 QTL for development-related traits, 23 QTL for morpho-physiological traits, 23 QTL for yield-related traits, 2 QTL for grain yield and 4 QTL for GPC (Table 5-8, Fig. 10). Adjusted R^2 of the CIM full model fitted for each trait is showed in Table 9. With regards to the subset of combined phenotypic data subjected to a comparative IBD-CIM/IBS-SMA QTL analysis, results are reported in Table 11-14 and Fig. 11. Considering the phenotypic data combined across years and environments IBD-CIM detected a total of 17 QTL effects with adjusted R^2 for the CIM full model decreasing from 0.52 (HD) and 0.40 (MD) to 0.26 (PH) and 0.12 (GY). The adjusted IBS-SMA identified 30 QTL for the four traits, including all but one of the 17 IBD-CIM QTL (Table 10, Fig. 10). The cumulative R^2 values of the QTL detected by IBS-SMA were almost equivalent (+/- 0.04) to the values obtained by IBD-CIM.

6.2 QTL mapping on single - environment data

Considering the IBD-CIM QTL analysis performed on single-environment data for HD, MD, PH and GY, a total of 23 QTL for Cad-11, 15 QTL for Cad-12, 13 QTL in Pr-11 and 9 QTL in Arg-12 (detailed results are presented in Table 11-14, Fig. 11) were identified. Comparing the results of QTL analysis for the combined-data with those obtained for single-environments, we found that 13 out of the 17 QTL detected with IBD-CIM across environments and 28 out of 30 QTL identified with the IBS-SMA showed environmental-specificity to various extents. In addition, several QTL with significant effects on single environments that were not significant across environments were found for all traits. IBD-CIM analysis for the single environments resulted in a total of 38 unique QTL (considering QTL with probability peaks within a 10-cM interval and consistency of direction of QTL effects across environments as unique QTL) for the four traits

considered. Of these, only four were consistently significant across all environments and were all detected for HD on chromosome 2A, 2B, 4B and 7A (Fig. 11) and were coincident with the main QTL clusters. Among the other QTL, two PH QTL were significant over three environments and corresponded to the QTL clusters on 2B proximal and 7A, 6 QTL including two for MD, two for PH and two for GY were concomitantly significant over two environments and the majority (26) were environment- and trait- specific. Among the 10 GY QTL, two QTL were concomitantly detected on chromosome 2A and 7A over two environments each, with the latest QTL corresponding to the 7A QTL cluster and the eight QTL remaining were all environment-specific.

6.3 QTL for development-related traits

Both CIM and SMA analysis identified four heading date QTL whose QTL peak positions were located with small (<5 cM) differences in genetic distances by the two mapping methods. These QTL, that showed the largest genetic effects on HD, were located on chr 2A (*QHd.ubo-2A*, with QTL peaks located at 46.0 and 42.0 cM with CIM and SMA and R^2 of 22.0 and 22.9%, respectively), chr 2B (*QHd.ubo-2B* at 51.5 and 57.0 cM, R^2 of 12.0 and 11.5%), chr 4B (at 59.1 and 55.1 cM, R^2 of 2.5 and 4.3%) and chr 7A (QHd.ubo-7A.2 at 63.2 and 65.7 cM, R^2 of 10.9% and 12.2%). At $P \leq 0.001$, IBS-SMA identified two additional HD QTL on chr. 6A (at 64.7 cM with R^2 of 3.4%) and on chr 7A (*QHd.ubo-7A.1* at 19.7 cM, R^2 of 5.3%). All the four main QTL on 2A, 2B, 4B and 7A showed concomitant significant effects across environments for maturity date, with magnitude and ranking of R^2 effects similar to that observed for HD. Additional MD QTL were identified by both CIM and SM analysis on chr 5B (peaks at 41.8 and 42.3 cM, respectively, with R^2 of 3.0 and 4.2%) and 7B (peaks at 25.0 identified by CIM only, R^2 of 1.3%); additional QTL effects for MD were found by SMA analysis only on chr 1B (two QTL), 1B, 2A, 2B, 4B, 6A (concomitant with the effect for HD as reported above), 7A, with R^2 ranging from 3.5 to 5.8% each. The analysis based on IBD-CIM allowed to test for the presence of different genetic effects among the four founders at each of the QTL. At the two major *QHd.ubo-2A* and *QHd.ubo-2B* QTL the post-hoc comparison showed that three of the four founder phenotypic effects were significantly diversified to each other. At *QHd.ubo-2A,* the latest heading phenotypic effect was associated to Colosseo allele (+3.05 days compared to the reference Rascon/Tarro), with Neodur allele associated to a delay of 1.94 days and Claudio and Rascon/Tarro with undistinguishable effects from each other that were categorized as the third earliest allele. At the *QHd.ubo-2A* CI region, the four founders' molecular haplotypes were easily distinguishable while at the most associated SNP the two alleles differentiated Colosseo and Neodur from Claudio and Rascon/Tarro (associated to an earliness effect of 2.69 days). At *QHd.ubo-2B*, the latest allele effect was associated to Neodur and Claudio (undistinguishable from each other, ca. +1.70 days from Rascon/Tarro), Rascon/Tarro was medium and Colosseo was the earliest (-1.28 days from

Rascon/Tarro). In this case, the local haplotype and the most associated SNP in the IBS-SMA effectively distinguished Colosseo and Neodur from Claudio and Rascon/Tarro while the distinguished features of Rascon/Tarro could not be detected. In the case of *QHd.ubo-7A* QTL, there were only two well distinct phenotypic effects, with Colosseo associated to earliness (-2.55 days from Rascon/Tarro) and distinct from all the other three founders. The QTL tag-SNP in the IBS-SMA distinguished Colosseo from the others founders and thus the two analysis yielded comparable results. Interestingly, Colosseo founder has distinct effects that were strong in magnitude at all the three major QTL, but with opposite directionality (i.e. contributing lateness at *QHd.ubo-2A* and earliness at *QHd.ubo-2B* at *QHd.ubo-7A*). Additional QTL for phenology showed a marked environment-specific expression and R^2 values ranging between approx. 1.5 and 5%. For the relatively stable QTL on chromosome 4B *QHd.ubo-4B* (detected in Cad11, Cad12, Pr11) three distinct phenotypic effects were associated to the four founders as already observed for group 2 QTL. Neodur and Colosseo founder alleles were both associated to a unique lateness effect (+0.5 days from Rascon/Tarro), Rascon/Tarro had a medium effect and Claudio's allele contributed an earliness effect (-0.76 days vs Rascon/Tarro). *QHd.ubo-4B* was most probably syntenic to the meta-QTL found by Hanocq *et al*. (2007) and to the A.18 GWAS-QTL found by LeGouis *et al*. (2012), as the QTL was positioned to a relative distance of 0.24, while the hexaploid wheat QTL were located at a relative distance of 0.31. In summary (and in respect of Rascon/Tarro effect), Neodur was always associated to lateness except for the QTL on 7A (*QHd.ubo-7A.2*) were the phenotypic effect was comparable to that of Rascon/Tarro, Claudio was associated to lateness at two QTL (2B and 7A) and to earliness at 2A and 4B QTL; Colosseo showed allelic effects opposite to those of Claudio, with lateness at 2A and 4B QTL and earliness at 2B and 7A QTL, Rascon/Tarro. For all the four QTL, the same trend have been reported for the corresponding MD QTL at target regions, with slight differences in relative effects. *QHd.ubo-2A*, *QHd.ubo-2B* and *QHd.ubo-7A* were confirmed to have concomitant effects on MD, however their effects were not as relevant as for HD, particularly for *QHd.ubo-2A* and *QHd.ubo-7A.* A range of QTL specific for MD and/or for MD and PH were also identified, all with low- to medium-R^2 values (*QMd.ubo-1B.1, QMd.ubo-1B.2, QMd.ubo-2A.2, QMd.ubo-2B.2, QMd.ubo-4B.2, QMd.ubo-5B, QMd.ubo-6A, QMd.ubo-7B*). Specifically for these QTL, Rascon/Tarro was frequently associated to early maturity, together with Claudio and/or Colosseo depending on the QTL (Claudio contributed a strong late maturity QTL at *QMd.ubo-7B* and Colosseo at *QMd.ubo-4B.2*). Neodur appeared to mostly contribute late maturity alleles.

6.4 QTL for morpho-physiological traits

QTL analysis for EGC identified three QTL, the most interesting one situated on chr 2A (R^2 = 8.8%) and co-mapped with a large NDVI_31 QTL (*QNdvi_31.ubo-2A*. R^2 = 16.6%) and a relatively lower-effect NDVI_40 QTL (*QNdvi_40.ubo-2A*. R^2 = 8.2%). Such QTL cluster is of physiological interest and appears to be related to yield-component and quality QTL (see discussion). Plant height data analysis on combined phenotypic data revealed five QTL confirmed by both IBD-CIM and IBS-SMA mapping methods, situated on chromosomes 1B (peak at 41.4 for mpMap and 38.3 for SMA, R^2 = 3.30% and 3.81%, respectively), 2B (peak at 97.9 and 96.8 cM, R^2 = 7.70% and 9.0%), 5B (peak at 105.8 and 105.3 cM, R^2 = 4.10 and 5.20%), 7A (peak at 67.2 cM for both methods, R^2 = 9.10 and 8.40%), 7B (138.4 and 152 cM, R^2 of and 2.8 and 3.5%). In addition, the SMA detected some other small effect quantitative loci for plant height on chr 2B (peak at 127.5, R^2 = 6.02%), 4A (peak at 120.0 cM, R^2 = 4.70%), 5B (peak at 105.3 cM, R^2 = 5.20%). Mapping of PH with single-environment data returned additionally QTL with relatively important effects in single environments: a PH QTL on chr 2A (peak at 42.5 cM, R^2 = 7.73%) with Cad11 data and a PH QTL on chr 5B (peak at 155.6 cM and R^2 = 6.53%) with Cad12 phenotypic dataset. The most significant locus associated to flag leaf chlorophyll content (SPAD index) was found on chr 5B (R^2 = 10.0%), co-localizing with a QTL for maturity date. They both appear not to drive any yield-component QTL.

6.5 QTL for yield components

Three QTL for FTsm have been found (chr 2A, chr 3A and chr 4B), although none of them explained a relative high portion of total genetic variation (R^2 = 3.9, 2.4 and 6.0%, respectively). Five QTL for SNE have been identified (chr 1B, chr 2A, chr 4B, chr 6B and chr 7A). The most important QTL mapped on chr 2A, showing an R^2 = 16.1% despite the relatively large confidence intervals (49.2 cM). Three SNE QTL (chr 1B with R^2 = 5.3%, chr 4B with R^2 = 9.9% and chr 6B with R^2 = 2.2%) were not found to be later represented in GNE. GNS showed a major QTL on chr 2A (R^2 = 39.2%), driven by a the positive effect of Rascon/Tarro haplotype which appears clearly functionally distinct by a second functional allele share by the other three founders. Interestingly, the GNS QTL located on chr 4B and belonging to a composite cluster appear not to be determined by any SNE QTL (see Discussion for further analysis of these QTL clusters). For GNE data, four significant QTL have been identified, the most interesting ones mapped on chr 2A (R^2 = 24.3%) and chr 7A (R^2 = 9.2%). Three QTL for GWE were identified, the main one on chr 7A (R^2 = 8.6%). Finally, four GWT QTL were identified, including a major locus on chr 2A which co-mapped with major QTL for SNE, GNE and GNS as mentioned above (see discussion for further analysis). Summary statistics for QTL for yield components are reported in Table 7.

6.6 QTL for grain yield

Based on the four-environment combined phenotypic data, QTL for GY were found by both QTL analysis methods (IBD-CIM and IBS-SMA) on chromosome 2B (*QGy.ubo-2B*. Peak at 103.4 and 127.5 cM, R^2 = 3.90 and 5.30%, respectively) and 7A (*QGy.ubo-7A.* Peak at 110.9 and 127.5 cM, R^2 = 7.0 and 7.2%) (Table 10). It should be noted that for *QGy.ubo-7A*, CIM SIM QTL analysis provided somewhat contrasting results in terms of QTL peak position (not shown). In order to be as inclusive as possible , we decided for this QTL only to report the QTL CI interval based on SIM. Interestingly, as reported above, these two QTL appear to be differently correlated with PH and development-related traits. *QGy.ubo-2B* was found to co-map with the *QPh.ubo-2B* while its association with phenology was milder, as only a very small effect MD QTL (*QMd.ubo-2B.2*, detected by IBS-SMA only) co-mapped. On the contrary, *QGy.ubo-7A* belong to a QTL cluster including QTL for HD, MD and PH and coincident with *TaFT-7A* chromosome position. For the *QGy.ubo-2B,* Colosseo allele was associated to a decrease of GY (-0.35 ton/ha) in respect to the other three founders, whose allelic effects were indistinguishable from each other. The directionality of allelic effects agreed with the effects observed for PH where Colosseo showed an effect for reduced PH. The QTL-tag SNP at this GY QTL in part failed to predict the founders' allelic effects because the two alleles related Claudio to Colosseo and Neodur to Rascon/Tarro, respectively. At *QGy.ubo-7A,* the post-hoc test differentiated three functional alleles, with Colosseo associated to a sensible decrease in GY (-0.35 ton/ha), Claudio also associated to a yield decrease (-0.11 ton/ha) in comparison with Neodur and Rascon/Tarro, which appear to share a similar effect QTL allele. At both chr 2B and 7A GY QTL the earliness allele of Colosseo seemed to be associated to a negative effect on yield.

Most of the GY QTL were detected for single-environment trials only. We identified eight GY QTL of which only one (identified in Cad11 and Arg12) corresponded to a GY QTL found with combined data (*QGy.ubo*-7A). Most significant single-environment GY QTL data were represented by QTL mapped with Cad11 data on chr 2A (42.51, R^2 = 5.69%) and 2B (51.54, R^2 = 4.84%) and by a QTL located at 101.2 cM on chr 4B (R^2 = 6.02%) with Arg12 data.

6.7 Quality traits

Regarding quality-related traits, we reported four QTL for GPC, of which two co-mapped with GWT (chr 2A, R^2 of 12.0% and chr 4B, R^2 of 5.6%) and two localized in the same genomic region of a GY QTL (chr 2B, R^2 of 3.2% and chr 7A, R^2 of 4.0%).

Table 5 Development-related traits. IBD-CIM analysis on combined phenotypic means. Threshold for QTL detection: *p*-value <10^{-3}; LOD-2 confidence interval).

QTL name	Pos	s.i.	R^2	*p*- value	LOD	L Mrk-R Mrk	Eff_Neo	Eff_Cla	Eff_Col
QHd.ubo -2A	46.0	41.5-49.5	22.0	0	>15.0	IWB70098 - IWB52303	1.94	-0.45	3.05
QHd.ubo -2B	51.5	39.5-53.5	12.0	0	>15.0	IWB4604 - IWA3868	1.6	1.73	-1.28
QHd.ubo -4B	59.1	49.1-70.1	2.5	1.42E-04	3.1	IWB6994 - IWB2398	0.6	-0.76	0.5
QHd.ubo -7A.2	63.2	62.2-67.2	10.9	0	>15.0	IWA8390 - IWB72200	0.15	0.45	-2.55
QMd.ubo- 2A.1	45.0	33.4-71.5	11.5	0	> 15.0	IWB67517 - IWB70098	0.69	-0.32	1.72
QMd.ubo- 2B.1	51.5	39.5-63.1	10.7	6.79E-14	12.2	IWB4604 - IWA3868	0.46	1.05	-1.1
QMd.ubo- 4B.2	97.2	85.7-108.2	3.9	4.10E-06	4.6	IWB67166 - IWB25207	0.49	0.64	1.16
QMd.ubo- 5B	41.8	30.8 -51.3	3.0	1.99E-05	4.0	IWA3870 - IWB45033	0.81	0.97	0.36
QMd.ubo- 7A	63.2	58.7 -67.2	4.9	2.89E-09	7.7	IWA8390 - IWB72200	0.32	0.55	-0.87
QMd.ubo- 7B	25.0	0 -48.6	1.3	6.84E-04	2.5	IWA2568 - IWA6901	0.27	1.46	0

Table 6 Morpho-physiological traits. IBD-CIM analysis on combined phenotypic means. Threshold for QTL detection: *p*-value <10^{-3}; LOD-2 confidence interval).

QTL name	Pos	s.i.	R^2	pvalue	LOD	L Mrk-R Mrk	Eff_Neo	Eff_Cla	Eff_Col
QEgc.ubo -2A	97.1	83.1 - 106.2	8.8	1.99E-08	6.8	IWB35615 - IWB34999	-0.14	0.2	0.43
QEgc.ubo -4B	75.2	64.6 - 85.7	4.3	1.87E-05	4.0	IWB68348 - IWB32941	-0.06	0.26	-0.23
QEgc.ubo -5A	82.1	70.6 - 93.8	3.6	3.91E-04	2.7	IWB14445 - IWB14333	0.21	0.03	-0.17
QFler.ubo -2A	42.5	6.3 - 71.5	3.0	5.37E-03	1.7	IWB51686 - IWB70375	-0.42	-0.27	-0.34
QFler.ubo -4B	40.6	29.1 - 50.6	5.6	9.81E-06	4.2	IWB46249 - IWB69708	-0.43	-0.56	-0.2
QFler.ubo -7B	95.8	85.8 - 137.9	3.2	7.69E-04	2.5	IWB72925 - IWB71827	-0.22	-1.22	-0.56
QPh.ubo- 1B	41.4	31.3 -51.9	3.3	6.20E-04	2.5	IWB66840 - IWB8804	2.92	0.94	1.64
QPh.ubo- 2B	97.9	87.8 -127.5	7.7	4.64E-08	6.5	IWB69270 - IWB69796	1.05	0.12	-2.91
QPh.ubo- 5B	105.8	95.7 -126.6	4.1	9.83E-05	3.3	IWB3693 - IWB71533	0.14	-2.48	0.15
QPh.ubo- 7A	67.2	60.7 -76.8	9.1	9.71E-08	6.2	IWB67995 - IWB33919	-0.5	0.4	-3.91
QPh.ubo- 7B.2	138.4	104.8 -180.5	2.8	7.87E-04	2.4	IWB10498 - IWA7330	-0.22	3.61	1.32
QFlg.ubo -2A	50	33.4 - 71.5	3.8	2.02E-05	3.9	IWB71058 - IWB67178	-0.53	-0.43	-1.87
QFlg.ubo -2B	127.5	103.9 - 137.5	5.2	7.24E-05	3.4	IWB43954 - IWB43015	-0.04	-0.82	-1.55
QFlg.ubo -3A	64.2	49.6 - 75.2	4.5	2.45E-05	3.9	IWB71974 - IWB70107	-0.63	-1.77	-1.22
QFlg.ubo -5B	42.3	32.8 - 51.3	10.0	1.31E-09	8.0	IWB45033 - IWB57461	-2.11	-0.47	0.11
QNdvi_31.ubo -2A	95.6	93.1 - 106.2	16.6	1.81E-13	11.8	IWB8192 - IWB9037	-0.02	0.01	0.03
QNdvi_31.ubo -6B	106.2	89.6 - 119.8	6.4	2.32E-04	2.9	IWB3431 - IWA221	-0.02	0.01	0.01
QNdvi_40.ubo -2A	95.6	84.6 - 106.2	8.2	3.77E-06	4.6	IWB8192 - IWB9037	0	0.01	0.03
QNdvi_40.ubo -6B	106.2	89.6 - 119.8	6.4	9.58E-05	3.3	IWB3431 - IWA221	-0.02	0	0.01
QNdvi_57.ubo -1A	27.6	2 - 82.6	3.0	3.07E-03	1.9	IWB3682 - IWB73652	0.01	0.01	0
QNdvi_57.ubo -2A	50.5	33.4 - 71.5	6.9	1.22E-06	5.1	IWB67178 - IWB53117	0.01	-0.01	0.02
QNdvi_57.ubo -2B	53.5	23.7 - 84.3	3.1	9.41E-04	2.4	IWA3868 - IWB30115	0.01	0.02	0
QNdvi_57.ubo -6B	106.2	89.6 - 131.4	5.9	3.61E-05	3.7	IWB3431 - IWA221	-0.01	0.01	0.01

Table 7 Yield and yield-related traits. IBD-CIM analysis on combined phenotypic means. Threshold for QTL detection: *p*-value <10^{-3}; LOD-2 confidence interval).

QTL name	Pos	s.i.	R^2	pvalue	LOD	L Mrk-R Mrk	Eff_Neo	Eff_Cla	Eff_Col
QFTsm.ubo -2A	109.2	98.1 - 128.9	3.9	1.42E-04	3.1	IWB28520 - IWB35417	-9.85	17.16	24.1
QFTsm.ubo -3A	109.8	84.7 - 128.4	2.4	5.40E-04	2.6	IWB71364 - IWB67653	-15.01	-40.63	-11.26
QFTsm.ubo -4B	95.7	85.7 - 101.2	6.0	1.20E-06	5.1	IWB72291 - IWB67166	13.34	3.1	-20.5
QSne.ubo -1B	132	124.2 - 148.7	5.3	1.13E-05	4.2	IWB10224 - IWB73738	-0.6	-0.38	-0.68
QSne.ubo -2A	81.1	33.4 - 82.6	16.1	0.00E+00	>15.0	IWB73849 - IWB20811	1.32	-0.39	0.5
QSne.ubo -4B	60.1	55.6 - 70.1	9.9	1.05E-11	10.0	IWB73588 - IWB11999	0.2	-0.29	0.72
QSne.ubo -6B	61.5	29 - 119.8	2.2	2.92E-03	1.9	IWB72280 - IWB42585	-0.15	0.02	0.34
QSne.ubo -7A	65.7	61.7 - 72.3	7.2	7.11E-12	10.2	IWB72200 - IWB67995	0.19	0.1	-0.84
QGne.ubo -2A	83.6	81.1 - 84.6	24.3	0.00E+00	>15.0	IWB32396 - IWB27892	-4.92	-10.6	-7.67
QGne.ubo -4B	116.2	106.2 - 126.2	3.0	7.94E-08	6.3	IWB68421 - IWB71580	3.43	1.17	4.33
QGne.ubo -6A	20.6	0 - 103.7	1.5	1.16E-03	2.3	IWB63758 - IWB9810	0.11	0.29	3.35
QGne.ubo -7A	60.7	58.7 - 67.2	9.2	1.30E-14	12.9	IWB73088 - IWB47104	2	1.76	-4.12
QGns.ubo -2A	86.1	84.6 - 87.6	39.2	0.00E+00	>15.0	IWA581 - IWB72980	-0.54	-0.58	-0.55
QGns.ubo -4B	129.7	118.2 - 139.2	1.7	5.78E-09	7.4	IWB5588 - IWB27042	0.22	0.17	0.23
QGns.ubo -7A	61.7	58.7 - 72.3	4.2	3.62E-08	6.6	IWB47104 - IWB8374	0.09	0.08	-0.13
QGwe.ubo -1B	38.3	24 - 48.4	7.0	4.45E-07	5.5	IWB73662 - IWB72569	0.18	-0.04	0.16
QGwe.ubo -2A	42.5	24.4 - 71.5	2.4	6.27E-05	3.5	IWB51686 - IWB70375	0.15	0.03	0.18
QGwe.ubo -4B	98.7	79.6 - 148.8	3.7	3.68E-04	2.8	IWB35646 - IWB7850	0.05	0.08	0.19
QGwe.ubo -7A	59.7	53 - 72.3	8.6	1.02E-09	8.1	IWB59817 - IWB73088	0.11	0.12	-0.17
QGwt.ubo -1B	30.3	24 - 48.9	5.6	1.09E-12	11.0	IWB9147 - IWB8046	3.38	0.16	3.08
QGwt.ubo -2A	83.6	82.6 - 84.6	35.7	0.00E+00	>15.0	IWB32396 - IWB27892	6.4	7.94	6.45
QGwt.ubo -4B	139.2	127.2 - 149.3	3.5	8.26E-07	5.3	IWB73977 - IWA3654	-2.71	-0.49	-1.58
QGwt.ubo -5A	37.6	23.3 - 51.6	2.6	1.10E-04	3.2	IWA1829 - IWB71757	-1.21	0.93	0.85
QGy.ubo - 2B	103.4	87.8 -127.5	3.9	6.32E-05	3.5	IWB69796 - IWA5411	-0.14	-0.05	-0.35
QGy.ubo - 7A	110.9	98.8 -123.4	7.0	2.18E-07	5.8	IWB319 - IWB23989	0.16	-0.11	-0.39

Table 8 Quality traits. IBD-CIM analysis on combined phenotypic means. Threshold for QTL detection: *p*-value <10^{-3}; LOD-2 confidence interval).

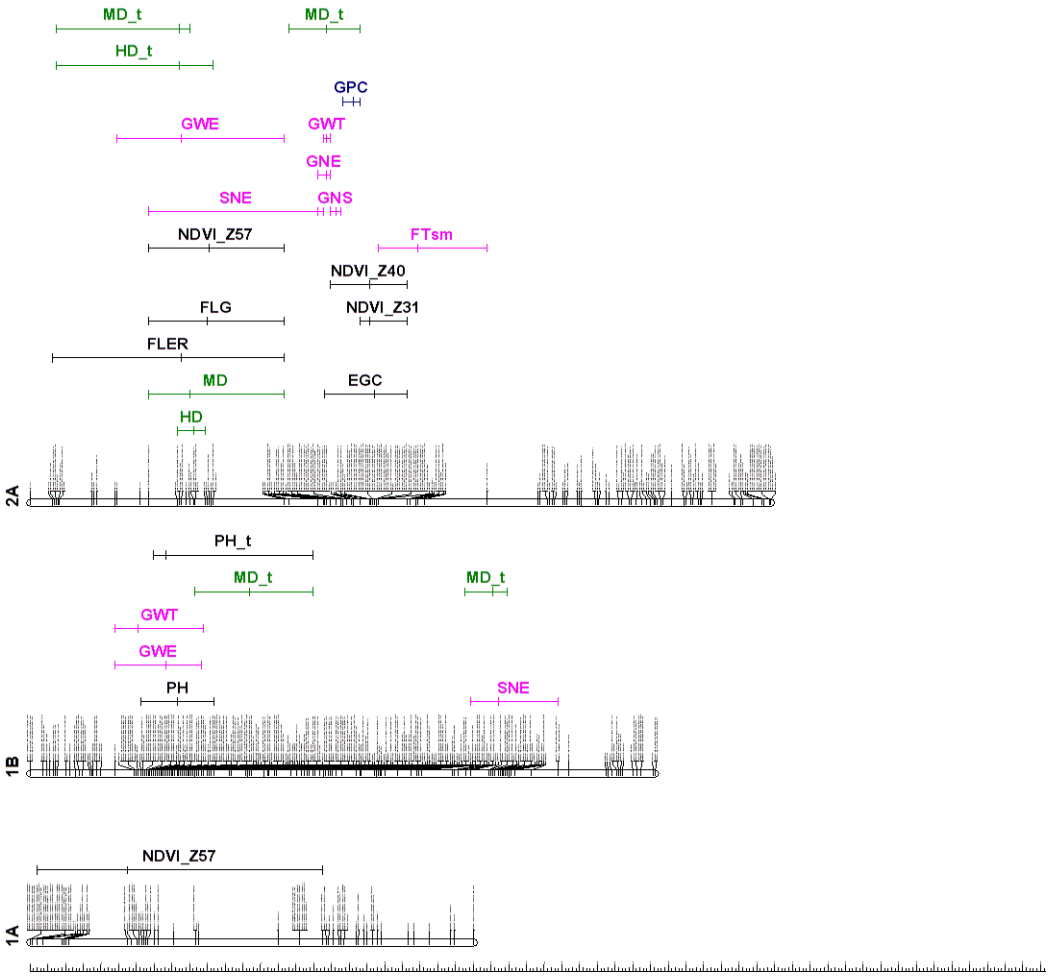
QTL name	Pos	s.i.	R^2	pvalue	LOD	L Mrk-R Mrk	Eff_Neo	Eff_Cla	Eff_Col
QGpc.ubo -1B	148.8	24 - 176.2	2.1	1.15E-04	3.2	IWA7992 - IWA3660	0.12	-0.29	0.16
QGpc.ubo -2A	91.1	88.1 - 93.1	12.0	2.86E-14	12.6	IWB32205 - IWB54818	0.48	0.82	0.65
QGpc.ubo -2B	136.5	84.3 - 172.6	3.2	8.39E-04	2.4	IWB32221 - IWB73241	-0.15	-0.26	0.15
QGpc.ubo -4B	129.7	118.2 - 140.7	5.6	1.43E-06	5.0	IWB5588 - IWB27042	-0.49	-0.2	-0.5
QGpc.ubo -7A	110.9	98.8 - 123.4	4.0	1.41E-05	4.1	IWB319 - IWB23989	-0.5	-0.14	0.08

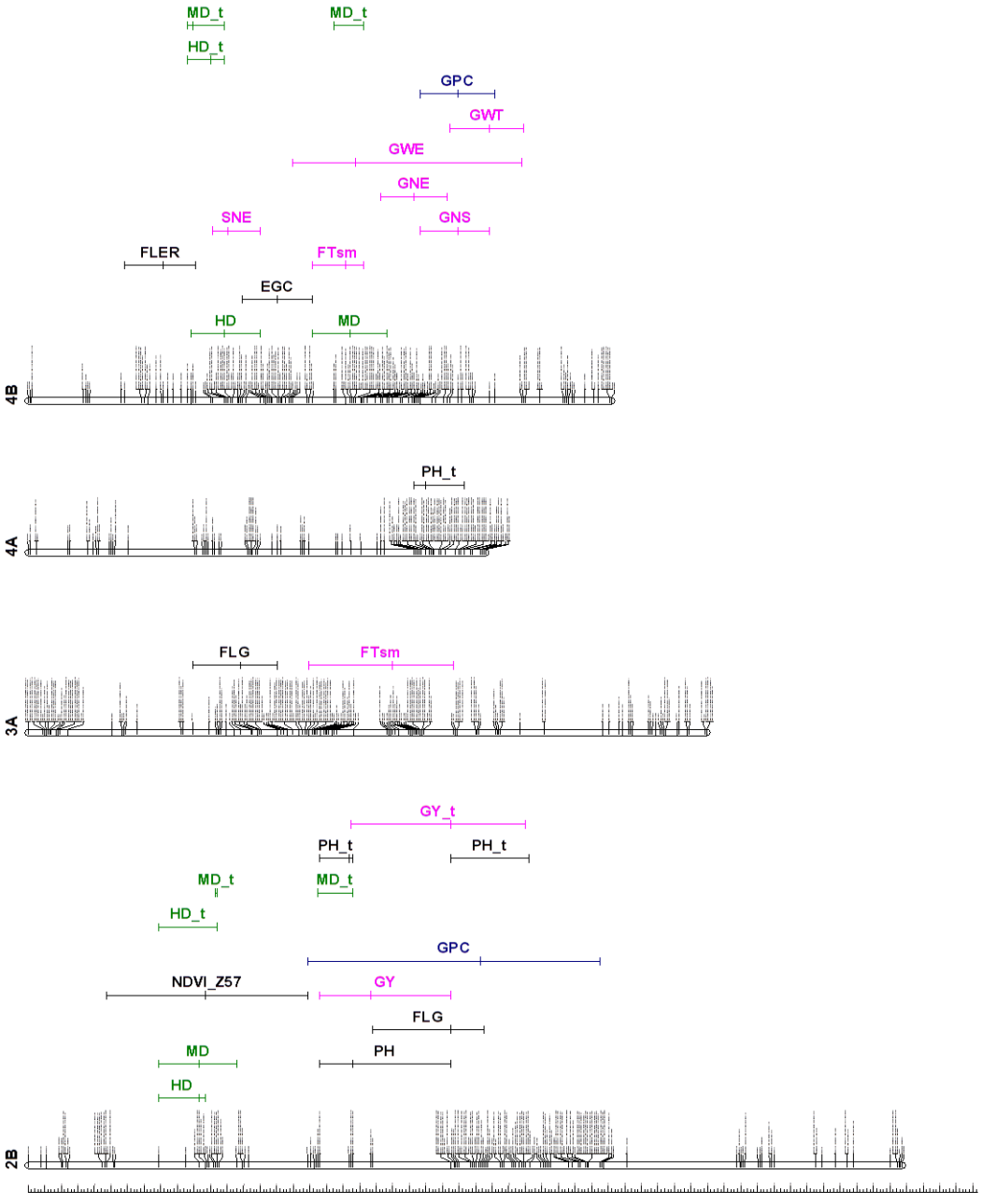
Table 9 The adjusted R^2 of the CIM full model for each of the trait which have undergone QTL analysis.

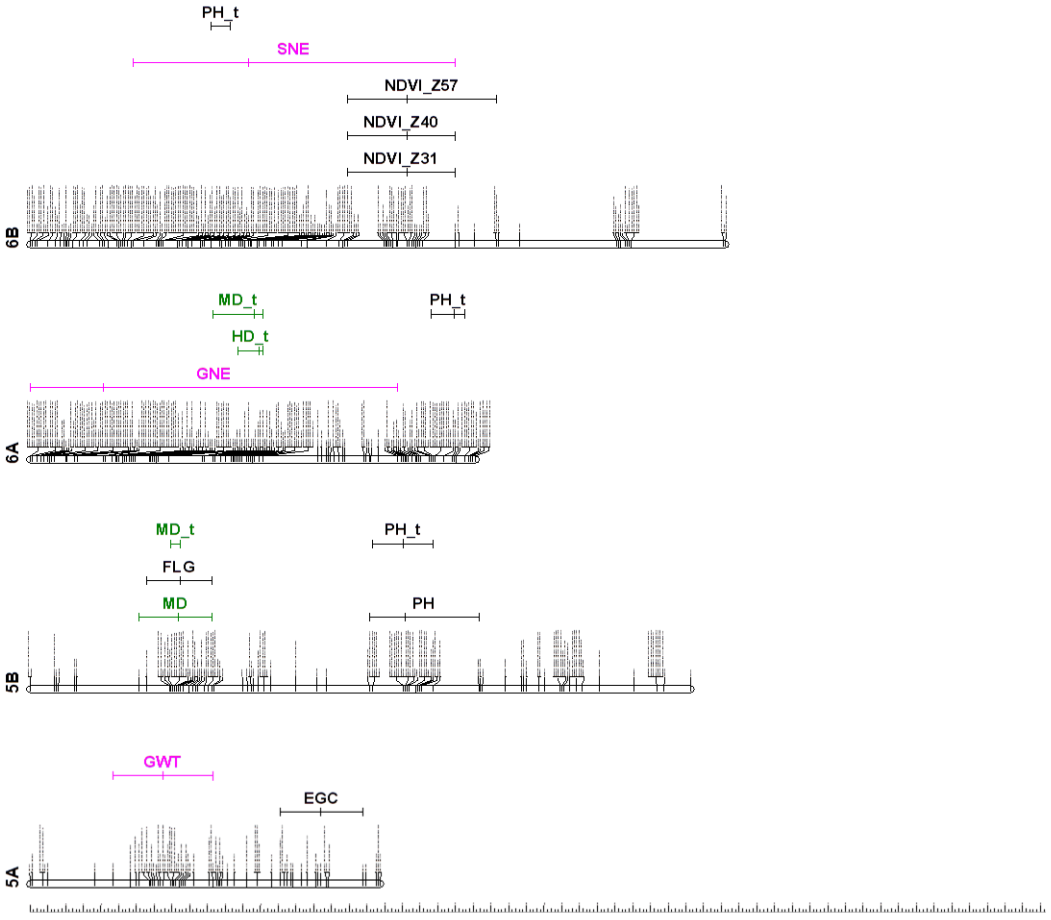
	IBD-CIM
Trait	**Adj. R^2 full model**
HD	0.52
MD	0.41
EGC	0.17
FLER	0.12
PH	0.26
FLG	0.24
NDVI_Z31	0.21
NDVI_Z40	0.13
NDVI_Z57	0.18
FTsm	0.14
SNE	0.41
GNE	0.43
GNS	0.50
GWE	0.23
GWT	0.50
GY	0.12
GPC	0.29

Table 10 Comparison between IBD-CIM and IBS-SMA methods for a subset of trait data (HD, MD, PH, GY).

	IBD-CIM										IBS-SMA							
HD	**Pos**	**s.i.**	**L Mrk-R Mrk**	***p*-value**	**R^2**	**LOD**	**Eff_Neo**	**Eff_Cla**	**Eff_Col**	**Funct. Alleles**	**Pos**	**Mrk**	**signif.i.**	**p-value**	**LOD**	**R^2**	**Effect**	**Ref**
QHd.ubo-2A	46.0	41.5-49.5	IWB70098 - IWB52303	0	22.0	> 15.0	1.94	-0.45	3.05	Col > Neo > Cla, RT	42	IWB67307	7.3 -51.5	3.32E-20	> 15.0	22.9	-2.69	Cla/RT
QHd.ubo-2B	51.5	39.5-53.5	IWB4604-IWA3868	0	12.0	> 15.0	1.6	1.73	-1.28	Neo, Cla > RT > Col	57	IWA8083	39.5 -57.1	2.43E-10	8.7	11.5	-1.90	Col/RT
QHd.ubo-4B	59.1	49.1-70.1	IWB6994-IWB2398	1.42E-04	2.5	> 15.0	0.6	-0.76	0.5	Neo, Col > RT > Cla	55.1	IWA4916	48.1 -59.1	1.37E-04	3.2	4.3	-1.17	Cla/RT
QHd.ubo-6A	-	-	-	-	-		-	-	-		64.7	IWA1235	58.7 -65.7	7.55E-04	2.5	3.4	-1.12	Cla/Col/RT
QHd.ubo-7A.1	-	-	-	-	-		-	-	-		19.7	IWB54744	0 -49	2.03E-05	3.9	5.3	1.52	Neo/Cla/RT
QHd.ubo-7A.2	63.2	62.2-67.2	IWA8390-IWB72200	0	10.9	> 15.0	0.15	0.45	-2.55	Neo, Cla, RT > Col	65.7	IWB56953	59.7 -98.8	6.06E-11	9.3	12.2	2.59	Neo/Cla/RT
MD	**Pos**	**s.i.**	**L Mrk-R Mrk**	**pvalue**	**R^2**	**LOD**	**Eff_Neo**	**Eff_Cla**	**Eff_Col**	**Funct. Alleles**	**Pos**	**Mrk**	**signif.i.**	**p-value**	**LOD**	**R^2**	**Effect**	**Ref**
QMd.ubo- 1B.1	-	-	-	-	-		-	-	-		61.8	IWB66172	56.6 -68.9	1.09E-05	4.2	5.7	-0.96	Cla/Col/RT
QMd.ubo- 1B.2	-	-	-	-	-		-	-	-		130.5	IWB6805	122.6 -134.5	2.11E-04	3.0	4.1	-0.75	Cla/RT
QMd.ubo- 2A.1	45.0	33.4-71.5	IWB67517-IWB70098	0	11.5	> 15.0	0.7	-0.3	1.72	Col > Neo > Cla, RT	42	IWB67305	7.3 -45	1.79E-09	7.9	10.5	1.19	Cla/RT
QMd.ubo- 2A.2	-	-	-	-	-		-	-	-		83.6	IWB1	73 -91	1.01E-05	4.2	5.8	-1.04	RT
QMd.ubo- 2B.1	51.5	39.5-63.1	IWB4604-IWA3868	6.8E-14	10.7	> 15.0	0.46	1.05	-1.1	Neo, Cla, RT > Col	57	IWA8083	56.5 -57.1	1.47E-08	7.0	9.3	-1.13	Col/RT
QMd.ubo- 2B.2	-	-	-	-	-		-	-	-		97.9	IWB5039	87.3 -97.9	7.48E-04	2.5	3.5	0.81	Neo/Cla/RT
QMd.ubo- 4B.1	-	-	-	-	-		-	-	-		49.6	IWB72203	48.11 -59.1	1.08E-05	4.2	5.7	-0.88	Cla/RT
QMd.ubo- 4B.2	97.2	85.7-108.2	IWB67166-IWB25207	4.10E-06	3.9	4.6	0.49	0.64	1.16	Neo, Cla, Col > RT	101.2	IWB66623	92.2 -101.2	3.24E-04	2.8	3.9	-0.78	RT
QMd.ubo- 5B	41.8	30.8 -51.3	IWA3870-IWB45033	1.99E-05	3.0	4.0	0.81	0.97	0.36	Neo, Cla, Col > RT	42.3	IWB37497	39.8 -42.3	3.03E-04	2.8	4.2	-0.76	Col/RT
QMd.ubo- 6A	-	-	-	-	-		-	-	-		63.2	IWA6724	51.7 -65.7	2.65E-04	2.9	3.9	-0.81	Cla/Col/RT
QMd.ubo- 7A.1	-	-	-	-	-		-	-	-		48.7	IWB71609	35.7 -49.7	8.04E-05	3.4	4.6	-0.8	Col/RT
QMd.ubo- 7A.2	63.2	58.7 -67.2	IWA8390-IWB72200	2.9E-09	4.9	7.7	0.32	0.55	-0.87	Neo, Cla, RT > Col	62.7	IWB40574	62.7 -67.2	2.36E-05	3.9	5.3	1.05	Neo/Cla/RT
QMd.ubo- 7B	25.0	0 -48.6	IWA2568-IWA6901	6.84E-04	1.3	2.5	0.27	1.46	0	Cla > Neo, Col, RT	-	-	-	-	-	-	-	-
PH	**Pos**	**s.i.**	**L Mrk-R Mrk**	**pvalue**	**R^2**	**LOD**	**Eff_Neo**	**Eff_Cla**	**Eff_Col**	**Funct. Alleles**	**Pos**	**Mrk**	**signif.i.**	**p-value**	**LOD**	**R^2**	**Effect**	**Ref**
QPh.ubo- 1B	41.4	31.3 -51.9	IWB66840-IWB8804	6.20E-04	3.3	2.5	2.92	0.94	1.64	Neo, Cla, Col > RT	38.3	IWB58817	34.8 -79.7	3.91E-04	2.7	3.8	-2.24	Cla/Col/RT
QPh.ubo- 2B	97.9	87.8 -127.5	IWB69270-IWB69796	4.6E-08	7.7	6.5	1.05	0.12	-2.91	Neo, Cla, RT > Col	96.8	IWB69109	87.7 -97.9	2.61E-08	6.7	9.0	3.53	Neo/Cla/RT
QPh.ubo- 2B	-	-	-	-	-		-	-	-		127.5	IWB20922	127.5 -151	6.28E-06	4.4	6.0	2.78	Neo/Cla/RT
QPh.ubo- 4A	-	-	-	-	-		-	-	-		120	IWB23377	116.5 -131.5	6.38E-05	3.5	4.7	-2.20	Neo/RT
QPh.ubo- 5B	105.8	95.7 -126.6	IWB3693-IWB71533	9.83E-05	4.1	3.3	0.14	-2.48	0.15	Neo, Col, RT > Cla	105.3	IWB11813	96.7 -113.8	2.56E-05	3.8	5.2	2.68	Neo/Col/RT
QPh.ubo- 6A	-	-	-	-	-		-	-	-		119.7	IWA6537	113.2 -122.7	2.67E-05	3.8	5.3	2.79	Neo/Cla/RT
QPh.ubo- 6B	-	-	-	-	-		-	-	-		51	IWA2451	51-56.5	8.18E-04	2.4	3.5	-2.30	Neo/Cla/RT
QPh.ubo- 7A	67.2	60.7 -76.8	IWB67995-IWB33919	9.7E-08	9.1	6.2	-0.5	0.4	-3.91	Neo, Cla, RT > Col	67.2	IWA7301	59.7 -98.8	1.46E-07	6.0	8.4	3.89	Neo/Cla/RT
QPh.ubo- 7B.1	-	-	-	-	-		-	-	-		14	IWB27108	14 -14	1.37E-04	3.2	4.3	-2.36	Cla/Col/RT
QPh.ubo- 7B.2	138.4	104.8 -180.5	IWB10498-IWA7330	7.87E-04	2.8	2.4	-0.22	3.61	1.32	Cla, Col > Neo, RT	152	IWB69205	138.4 -158	5.24E-04	2.6	3.5	2.24	Cla/Col/RT
GY	**Pos**	**s.i.**	**L Mrk-R Mrk**	**pvalue**	**R^2**	**LOD**	**Eff_Neo**	**Eff_Cla**	**Eff_Col**	**Funct. Alleles**	**Pos**	**Mrk**	**signif.i.**	**p-value**	**LOD**	**R^2**	**Effect**	**Ref**
QGy.ubo- 2B	103.4	87.8 -127.5	IWB69796IWA5411	6.32E-05	3.9	3.5	-0.14	-0.05	-0.35	Neo, Cla, RT > Col	127.5	IWB62437	97.4 -150	3.08E-05	3.8	5.3	0.23	Neo/RT
QGy.ubo- 7A	110.9	98.8 -123.4	IWB319IWB23989	2.18E-07	7.0	5.8	0.16	-0.11	-0.39	Neo > Cla, RT > Col	60.7	IWB30443	59.7 -87.8	6.81E-07	5.4	7.2	0.32	Neo/Cla/RT


MD_t
MD_t
HD_t
GPC
GWE
GWT
GNE
SNE
GNS
NDVI_Z57
FTsm
NDVI_Z40
FLG
NDVI_Z31
FLER
MD
EGC
HD
2A
PH_t
MD_t
MD_t
GWT
GWE
PH
SNE
1B
NDVI_Z57
1A


MD_t
MD_t
HD_t
GPC
GWT
GWE
GNE
SNE
GNS
FLER
FTsm
EGC
HD
MD
4B
PH_t
4A
FLG
FTsm
3A
GY_t
PH_t
PH_t
MD_t
MD_t
HD_t
GPC
NDVI_Z57
GY
FLG
MD
PH
HD
2B


PH_t
SNE
NDVI_Z57
NDVI_Z40
NDVI_Z31
6B
MD_t
PH_t
HD_t
GNE
6A
MD_t
PH_t
FLG
MD
PH
5B
GWT
EGC
5A

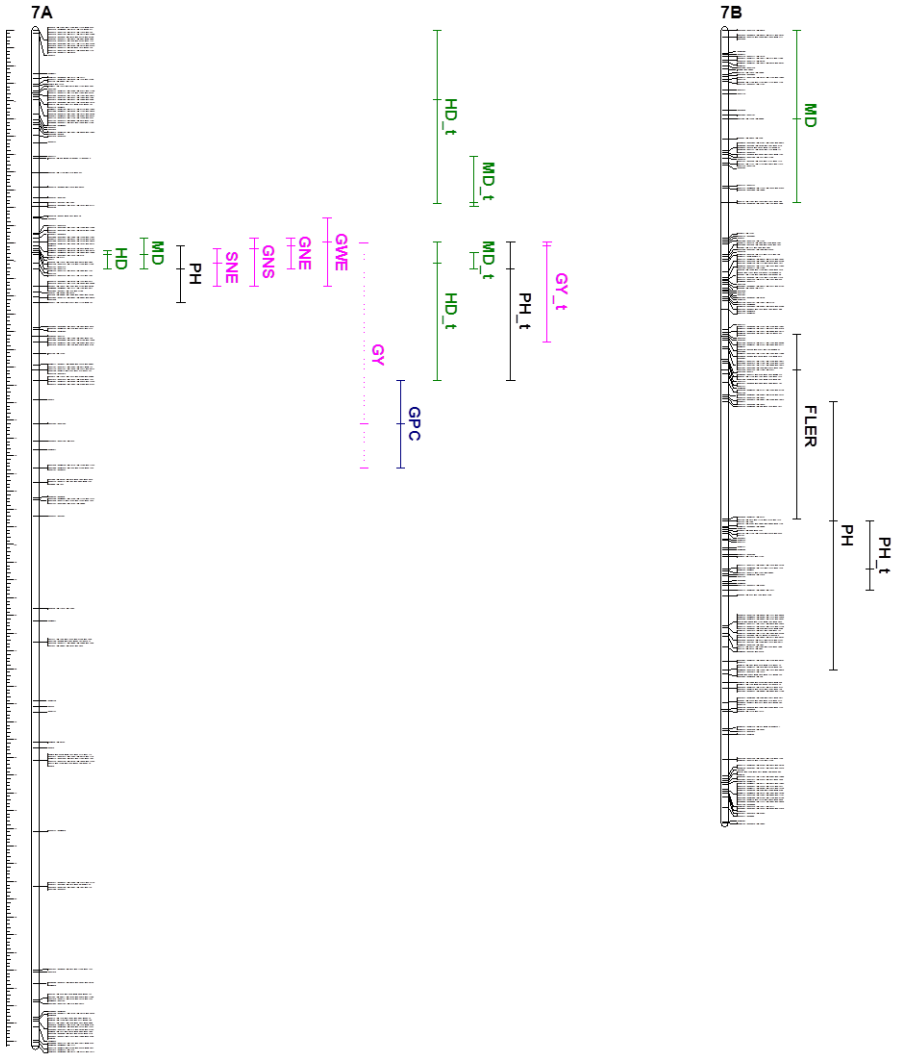


Fig 10 The position on the NCCR linkage map of the QTL detected for the considered traits (threshold for QTL detection: *p*-value $<10^{-3}$). For each QTL, the peak's position and the LOD-2 confidence interval are reported. IBS-SMA QTL confidence intervals have been calculated by a sliding window applied to SMA LOD values (± 4 marker, threshold at LOD equal to 1.44). Trait classes have been represented by different QTL bar colors: green, development-related traits; black, morpho-physiological traits; pink, yield and yield-related traits and blue, quality traits. The results of the IBS-SMA QTL analysis performed with Tassel for HD, MD, PH and GY are denoted as a t-HD, t-MD, t-PH and t-GY. Chromosomes and QTL have been drawn with MapChart 2.2 (Voorips, 2002).

Table 11 Single-environment Cad11 QTL detected by IBD-CIM (threshold for QTL detection: *p*-value < 10^{-3}; LOD-1 confidence interval).

Trait	Chr	Pos	s.i.	*p*- value	R^2	LOD	L Mrk - R Mrk	Eff_Neo	Eff_Cla	Eff_Col
HD_Cad11	1A	36.1	28.6 -125.2	2.57E-03	0.8	2.0	IWB63226 - IWA2287	1.06	0.35	-0.14
HD_Cad11	2A	46.0	33.4 -49.5	0	32.9	> 15.0	IWB70098 - IWB52303	3.15	-0.41	4.18
HD_Cad11	2B	51.5	39.5 -63.1	0	10.4	> 15.0	IWB4604 - IWA3868	1.8	1.84	-1.22
HD_Cad11	4B	59.1	49.1 -70.1	8.04E-04	2.7	2.4	IWB6994 - IWB2398	0.47	-0.85	0.47
HD_Cad11	7A	65.7	62.2 -67.2	0	9.0	> 15.0	IWB72200 - IWB67995	-0.03	0.2	-3.02
MD_Cad11	1A	40.6	30.6 -70	6.17E-05	2.0	3.5	IWA2287 - IWB51724	1.38	-0.53	0.38
MD_Cad11	1B	133.5	120.6 -176.2	1.43E-02	3.6	1.3	IWB9182 - IWB10582	1.11	0.09	0.84
MD_Cad11	2A	45.0	33.4 -71.5	0	16.7	> 15.0	IWB67517 - IWB70098	1.94	-0.72	3.16
MD_Cad11	2B	51.5	47.5 -63.1	8.27E-13	9.6	11.1	IWB4604 - IWA3868	0.5	1.58	-2
MD_Cad11	3A	76.2	25.3 -114.8	1.53E-03	1.6	2.2	IWB68422 - IWB71668	0.78	1.77	0.67
MD_Cad11	4B	97.7	92.7 -106.2	4.9E-09	4.4	7.4	IWB25207 - IWB71823	0.52	0.74	2.3
MD_Cad11	5B	42.3	8 -64.4	1.75E-03	2.3	2.1	IWB45033 - IWB57461	0.99	0.92	-0.16
MD_Cad11	7A	63.2	58.7 -72.3	4.07E-09	3.0	7.5	IWA8390 - IWB72200	0.62	1.05	-1.3
PH_Cad11	2A	42.5	33.4 -71.5	1.21E-13	7.7	11.9	IWB51686 - IWB70375	3.63	-0.27	3.87
PH_Cad11	2B	97.9	87.8 -127.5	2.16E-08	9.3	6.8	IWB69270 - IWB69796	0.76	0.13	-3.04
PH_Cad11	4A	75.2	58.2 -82.3	0.00000179	4.5	5.0	IWB48435 - IWB68425	-0.7	-1.1	2.96
PH_Cad11	6A	87.2	65.7 -98.2	1.79E-05	2.0	4.0	IWB72745 - IWB9468	-1.25	2.18	-2.17
PH_Cad11	6B	110.7	83.6 -196	7.98E-04	2.7	2.4	IWA634 - IWB47763	-2.67	-0.93	0.31
PH_Cad11	7A	63.2	61.7 -72.3	3.21E-14	10.7	12.5	IWA8390 - IWB72200	0.08	1.68	-3.91
GY_Cad11	2A	42.5	30.9 -51.5	0.000000139	5.7	6.0	IWB51686 - IWB70375	0.48	-0.02	0.37
GY_Cad11	2B	51.5	39.5 -63.1	1.49E-04	4.8	3.1	IWB4604 - IWA3868	0.38	0.45	0.08
GY_Cad11	3A	142.0	129.4 -155.6	3.94E-04	3.8	2.7	IWB20248 - IWB6370	-0.48	-0.24	-0.44
GY_Cad11	7A	63.2	58.7 -72.3	0.000000206	5.3	5.9	IWA8390 - IWB72200	0.26	0.24	-0.31

Table 12 Single-environment Cad12 QTL detected by IBD-CIM (threshold for QTL detection: *p*-value < 10^{-3}; LOD-1 confidence interval).

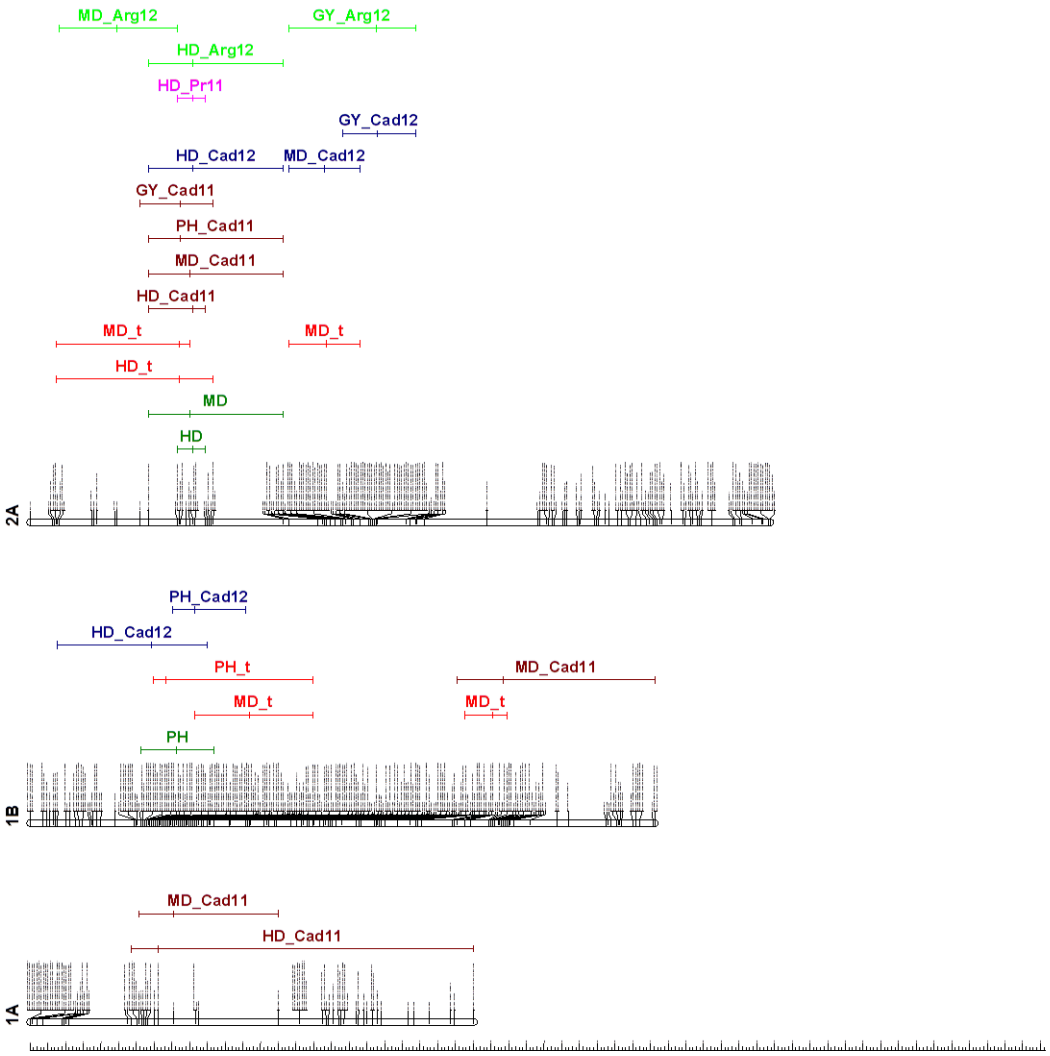
Trait	Chr	Pos	s.i.	*p*- value	R^2	LOD	L Mrk - R Mrk	Eff_Neo	Eff_Cla	Eff_Col
HD_Cad12	1B	34.3	7.7 - 49.9	1.99E-04	1.5	3.0	IWB39074 - IWB48416	0.4	-0.46	1.01
HD_Cad12	2A	46.0	33.4 - 71.5	1.11E-16	9.3	14.9	IWB70098 - IWB52303	1.32	-0.41	2.33
HD_Cad12	2B	51.5	39.5 - 63.1	1.39E-12	8.6	10.9	IWB4604 - IWA3868	1.74	1.71	-0.6
HD_Cad12	4B	55.1	43.6 - 65.6	1.13E-04	5.2	3.2	IWA4916 - IWB73830	1.12	-0.45	0.52
HD_Cad12	7A	63.2	62.2 - 65.7	0.00E+00	20.1	> 15.0	IWA8390 - IWB72200	-0.39	0.51	-3.62
MD_Cad12	2A	83.1	73 - 93.1	2.96E-07	7.0	5.7	IWB69369 - IWB32396	0.68	0.27	1.47
MD_Cad12	6A	58.7	36.1 - 109.2	7.85E-04	2.8	2.4	IWB66638 - IWB73301	0.64	0.18	0.73
MD_Cad12	7A	63.2	53 - 73.8	1.49E-05	5.0	4.1	IWA8390 - IWB72200	0.38	0.53	-0.45
PH_Cad12	1B	46.4	40.3 - 60.8	4.36E-04	3.3	2.7	IWB68449 - IWB7028	3.03	0.66	1.36
PH_Cad12	2B	97.9	66.6 - 151.5	1.09E-03	3.7	2.3	IWB69270 - IWB69796	1.1	0.16	-1.9
PH_Cad12	5B	155.6	145.1 - 170.3	1.83E-05	6.5	4.0	IWB44136 - IWB69060	-6.93	-14	-6.59
PH_Cad12	6A	94.7	85.7 - 104.7	1.15E-05	4.5	4.2	IWB65466 - IWB68272	-10.13	-12.96	-9.15
PH_Cad12	7A	67.2	58.7 - 83.8	6.22E-05	4.9	3.5	IWB67995 - IWB33919	-0.16	0.23	-3.46
PH_Cad12	7B	152.4	104.8 - 155.4	2.51E-05	3.7	3.9	IWB25127 - IWB71692	-1.54	2.28	1.7
GY_Cad12	2A	98.1	88.1 - 108.7	0.000428	4.3	2.7	IWB43067 - IWA32	0.62	0.55	0.41

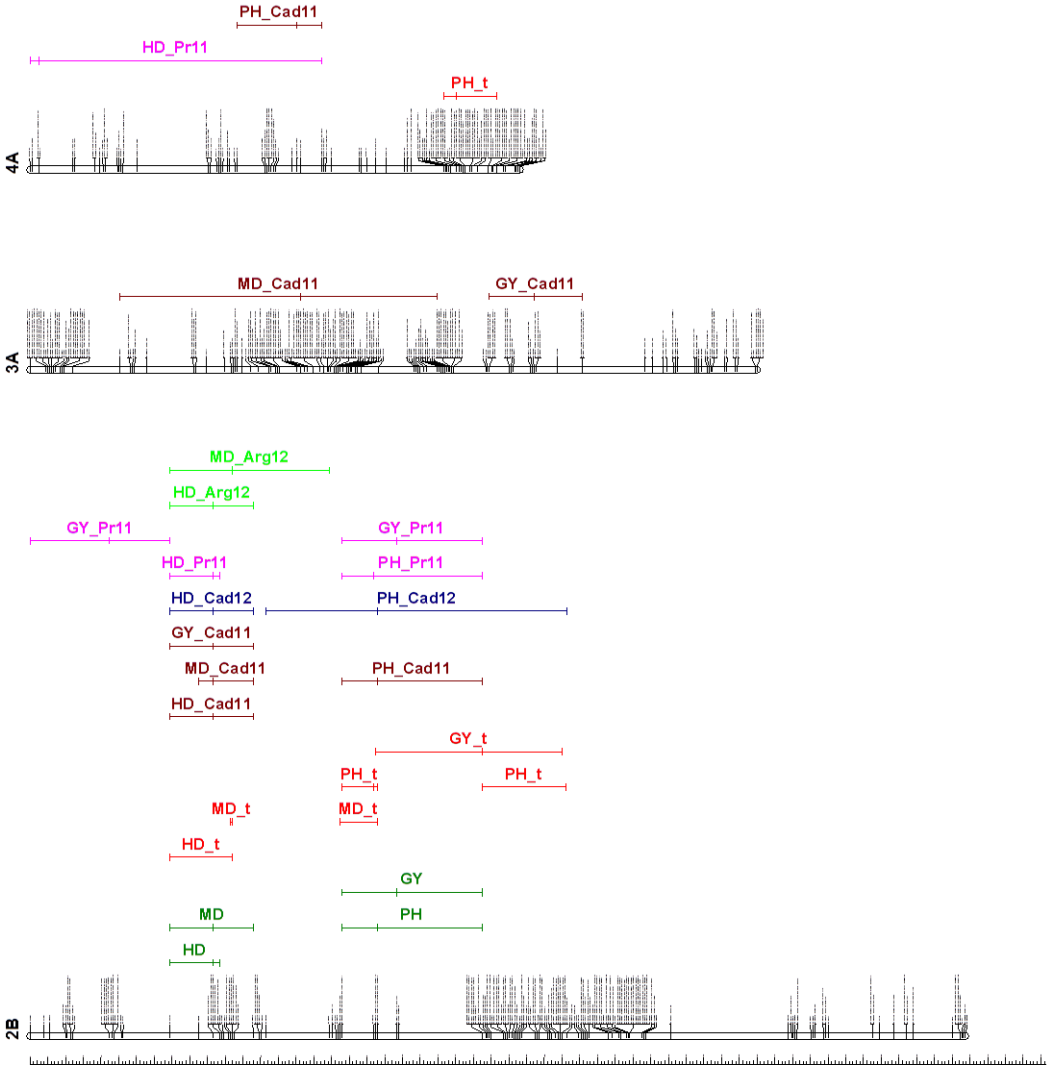
Table 13 Single-environment Pr11 QTL detected by IBD-CIM (threshold for QTL detection: *p*-value < 10^{-3}; LOD-1 confidence interval).

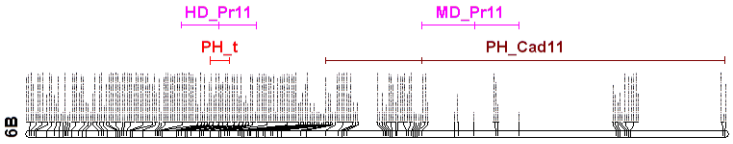
Trait	Chr	Pos	s.i.	*p*- value	R^2	LOD	L Mrk - R Mrk	Eff_Neo	Eff_Cla	Eff_Col
HD_Pr11	2A	46.01	41.5 -49.5	0.00E+00	30.2	>15.0	IWB70098 - IWB52303	2.7	-0.11	3.9
HD_Pr11	2B	51.54	39.5 -53.5	0.00E+00	10.8	>15.0	IWB4604 - IWA3868	1.65	1.84	-1.12
HD_Pr11	4A	2.5	0 -82.3	6.14E-04	1.8	2.5	IWB74057 - IWB26513	-0.61	-5.3	3.68
HD_Pr11	4B	59.14	49.1 -70.1	1.93E-04	2.9	3.0	IWB6994 - IWB2398	0.58	-0.68	0.63
HD_Pr11	6B	53.54	43 -64	2.52E-04	1.6	2.9	IWB59925 - IWA4078	0.18	0.22	1.19
HD_Pr11	7A	63.24	61.7 -67.2	4.44E-16	4.4	14.3	IWA8390 - IWB72200	0.53	0.83	-1.73
PH_Pr11	2B	96.87	87.8 -127.5	4.98E-07	5.7	5.5	IWB69363 - IWB69109	-0.36	0.16	-4.01
PH_Pr11	4B	28.09	18 -38.6	6.55E-04	2.7	2.5	IWB13072 - IWB63894	-2.54	-3.56	-1.44
PH_Pr11	6B	125.35	110.7 -137.9	3.14E-04	3.9	2.8	IWB71650 - IWA8072	-3.51	-0.31	-0.58
PH_Pr11	7A	67.24	65.7 -73.8	4.16E-05	6.6	3.6	IWB67995 - IWB33919	-0.37	1.09	-3.3
GY_Pr11	2B	22.16	0 -39.5	1.57E-04	4.2	3.1	IWB7031 - IWB36011	0.37	0.04	-0.01
GY_Pr11	2B	103.39	87.8 -127.5	6.49E-05	3.4	3.5	IWB69796 - IWA5411	-0.11	0.16	-0.31
GY_Pr11	7A	172.63	137 -189.2	1.28E-04	4.4	3.2	IWB68677 - IWA2371	0.36	-0.18	-0.18

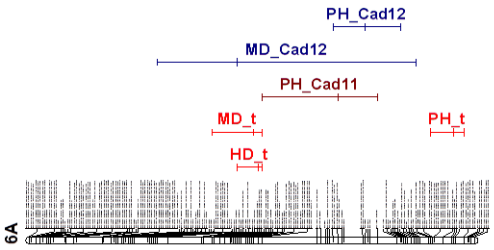
Table 14 Single-environment Arg12 QTL detected by IBD-CIM (threshold for QTL detection: *p*-value < 10^{-3}; LOD-1 confidence interval).

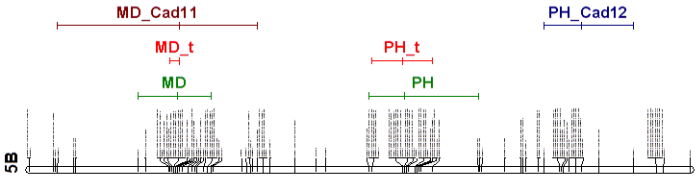
Trait	Chr	Pos	s.i.	*p*- value	R^2	LOD	L Mrk - R Mrk	Eff_Neo	Eff_Cla	Eff_Col
HD_Arg12	2A	46.0	33.4 -71.5	6.38E-13	8.7	11.2	IWB70098 - IWB52303	0.85	-0.76	1.9
HD_Arg12	2B	51.5	39.5 -63.1	8.90E-14	11.6	12.1	IWB4604 - IWA3868	1.05	1.17	-1.99
HD_Arg12	4B	118.2	108.2 -129.7	3.62E-03	1.6	1.8	IWB23828 - IWB3229	0.47	-0.79	0.2
HD_Arg12	7A	63.2	61.7 -72.3	1.00E-12	8.5	11.0	IWA8390 - IWB72200	0.43	0.16	-2.18
MD_Arg12	2A	24.4	8.3 -41.5	1.54E-04	5.2	3.1	IWB12622 - IWB71223	0.6	0.41	1.64
GY_Arg12	2B	57.0	39.5 -84.3	5.76E-05	5.7	3.5	IWB73972 - IWB46810	0.4	0.93	-0.79
GY_Arg12	2A	97.6	73 -108.7	4.97E-04	2.6	2.6	IWB34999 - IWB43067	-0.45	-0.39	-0.33
GY_Arg12	4B	101.2	92.7 -112.2	9.22E-07	6.0	5.2	IWB72802 - IWB73462	0.53	0.03	0.17
GY_Arg12	7A	62.2	49.7 -65.7	1.30E-05	4.9	4.1	IWB8374 - IWB117	0.17	0.03	-0.35

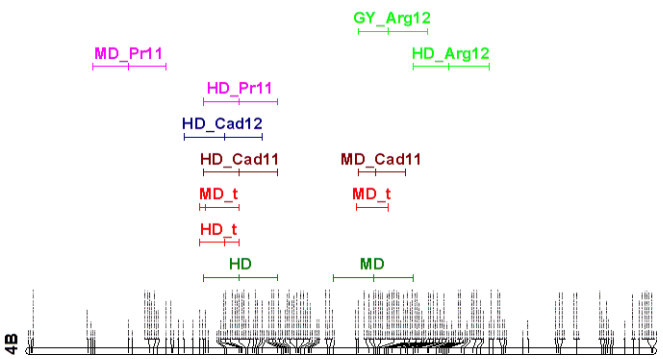

MD_Arg12
GY_Arg12
HD_Arg12
HD_Pr11
GY_Cad12
HD_Cad12
MD_Cad12
GY_Cad11
PH_Cad11
MD_Cad11
HD_Cad11
MD_t
MD_t
HD_t
MD
HD
2A
PH_Cad12
HD_Cad12
PH_t
MD_Cad11
MD_t
MD_t
PH
1B
MD_Cad11
HD_Cad11
1A


PH_Cad11
HD_Pr11
PH_t
4A
MD_Cad11
GY_Cad11
3A
MD_Arg12
HD_Arg12
GY_Pr11
GY_Pr11
HD_Pr11
PH_Pr11
HD_Cad12
PH_Cad12
GY_Cad11
MD_Cad11
PH_Cad11
HD_Cad11
GY_t
PH_t
PH_t
MD_t
MD_t
HD_t
GY
MD
PH
HD
2B


HD_Pr11
MD_Pr11
PH_t
PH_Cad11
6B


PH_Cad12
MD_Cad12
PH_Cad11
MD_t
PH_t
HD_t
6A


MD_Cad11
PH_Cad12
MD_t
PH_t
MD
PH
5B


GY_Arg12
MD_Pr11
HD_Arg12
HD_Pr11
HD_Cad12
HD_Cad11
MD_Cad11
MD_t
MD_t
HD_t
HD
MD
4B

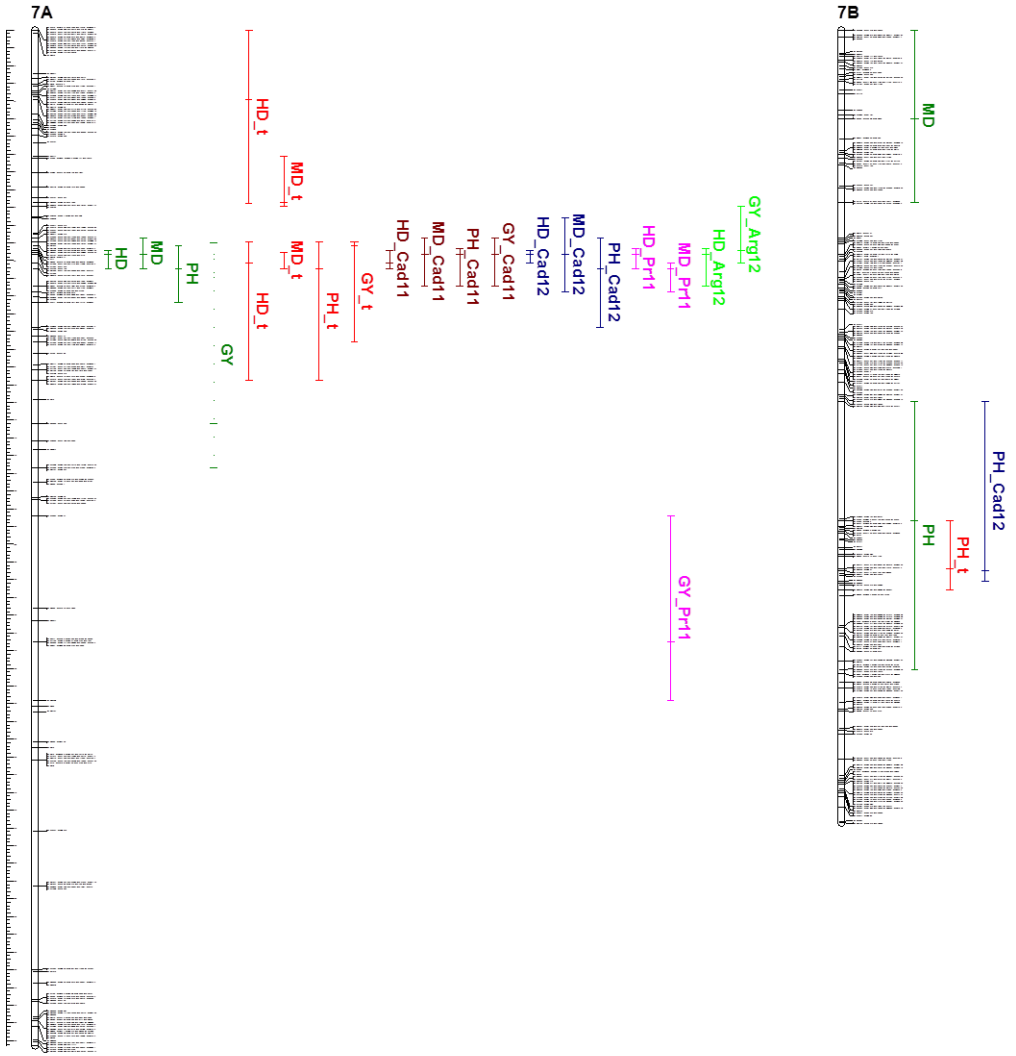


Fig 11 QTL results for the subset of trait data (HD, MD, PH and GY) which have been analysed both with IBS-SMA (red QTL bars) and IBD-CIM (green QTL bars). For each QTL, the peak's position and the LOD-2 confidence interval are reported (threshold for QTL detection: *p*-value $<10^{-3}$). The results of the QTL mapping on single-environment data are represented as following: brown, Cad11; blue, Cad12; pink, Pr11 and pale green, Arg12. For each QTL, the peak's position and the LOD-1 confidence interval are reported (threshold for QTL detection: *p*-value $<10^{-3}$). Chromosomes and QTL have been drawn with MapChart 2.2 (Voorips, 2002).

7. Identity of functional QTL haplotypes and comparison with SNP-based haplotypes at QTL

For the 17 QTL mapped across environments we verified the identity of functional haplotypes as obtained by CIM (i.e. based on estimated QTL effects for the four founders) with molecular haplotypes (i.e classification of founders haplotype based on SNP genotypes at QTL confidence interval). Despite the NCCR population is based on four founders, none QTL showed four functionally different haplotypes, while seven QTL showed three and ten showed two (mean number of functional QTL haplotypes per QTL = 2.4, see Table 10). At the same QTL, we identified four, three and two molecular haplotypes at 11, 5 and 1 QTL, respectively (mean value = 3.6 molecular haplotypes per QTL). Molecular haplotypes outnumbered functional haplotypes in all cases except two. At no QTL, the two grouping criteria provided the same results although some concordance was observed (e.g. extreme functionally different haplotypes were always classified as molecularly different except one case). Fig. 12 shows haplotype grouping at two exemplary case QTL, *QHd.ubo-2A* and *QHd.ubo-2B*. At *QHd.ubo-2A*, functional classification identified two haplotypes (Neodur = Colosseo, increasing heading date; Claudio = Rascon/Tarro decreasing heading date) while four haplotypes were identified based on SNP profiling (Fig. 12). In this case, functional and molecular haplotyping are not necessarily in contrast since the four molecular haplotypes may include only two functionally different alleles which are not obviously evident due to high genetic diversity between the four founders at this QTL. Indeed, shorter portion of this QTL confidence interval (as also outlined by the SNP-based QTL analysis - marker IWB67307. Fig. 12 indicates the presence of molecular haplotypes fully concordant with functional haplotypes. At *QHd.ubo-2B* (data not shown) functional haplotypes were three (Neodur = Claudio as late flowering, Rascon/Tarro as intermediate, and Colosseo as early flowering). However, SNP-based haplotypes were only two (Neodur *vs* Claudio, Colosseo and Rascon/Tarro). In this case, Claudio and Colosseo while attributed with the same molecular haplotype, showed contrasted functional effects. One possible explanation is that the relevant genetic variation (e.g. additional SNPs) linked with the functional gene(s) underlying this QTL was missed by our SNP profiling (see discussion).

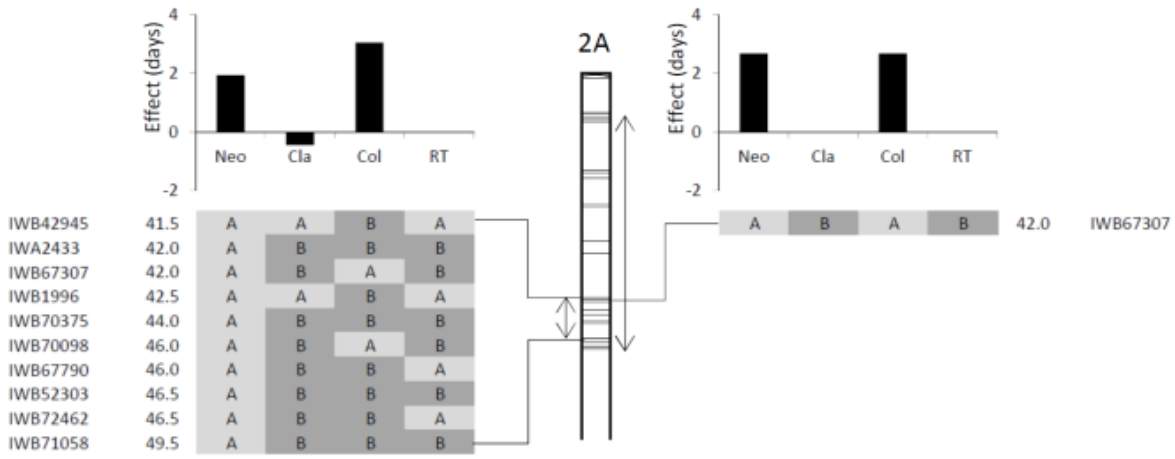


Fig 12 Functional and molecular haplotypes at *QHd.ubo-2A.*

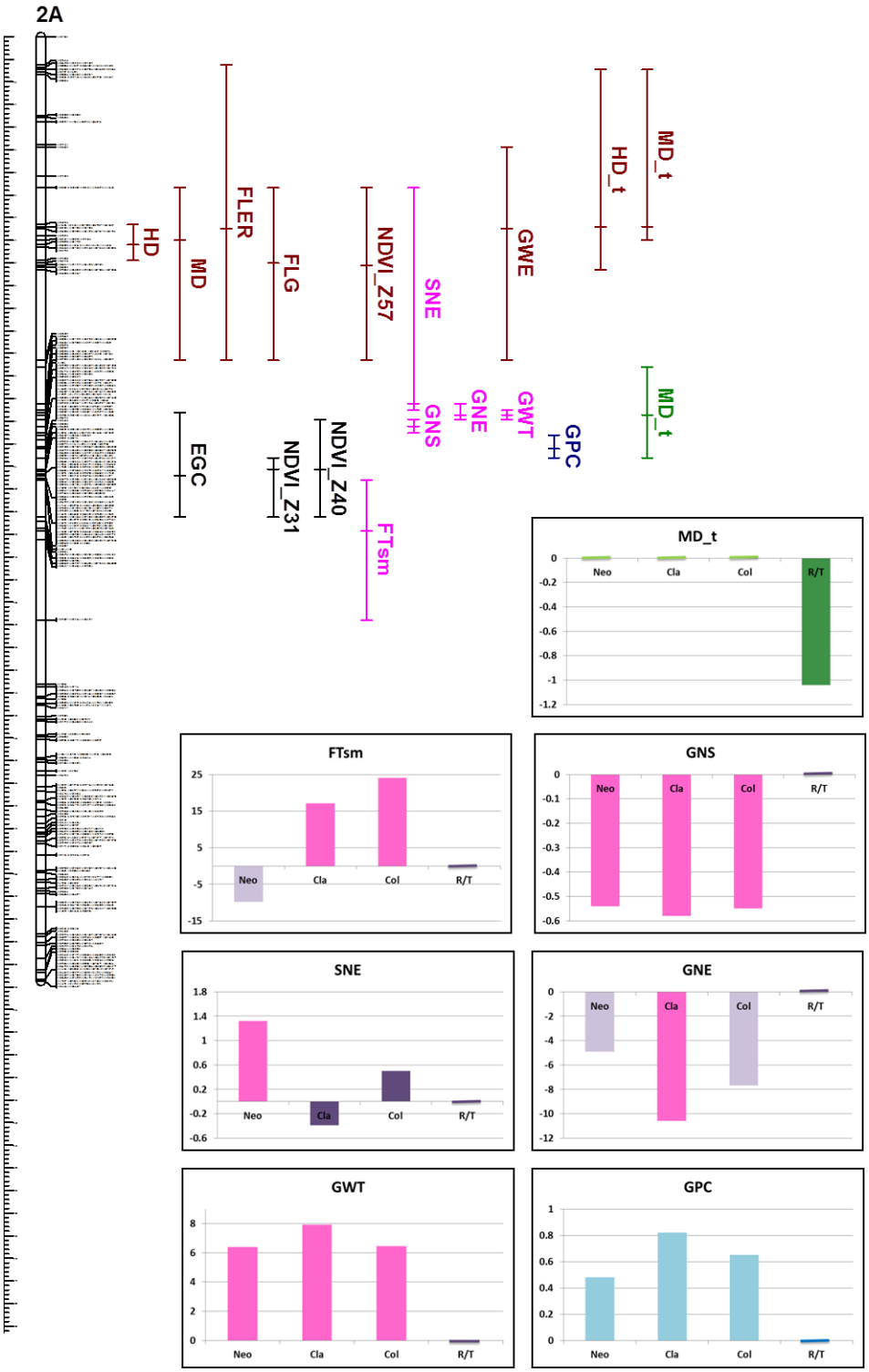


Fig. 13 The pattern of haplotype effects at the proximal QTL cluster on chr 2A. Trait in brown color are believed as not involved in the present cluster. The histograms represent the genetic effects estimated in terms of the four founders for MD, FTsm, GNS, SNE, GNE, GWT and GPC at the target region. The MD QTL was detected by IBS-SIM and therefore only the direction of the allelic effects could be estimated. Regarding the QTL for the other examined traits, they were identified by IBD-CIM. This enabled us to describe in the histograms the modulation of effects in terms of the four haplotypes (different color tones identify different functional haplotypes at each quantitative locus).

DISCUSSION

Increasing grain yield potential and stability and improving semolina quality are the major goals in durum wheat plant breeding (Slafer and Calderini, 2005). However the genetic basis of these traits is complex and its dissection remains a formidable task. Multiparental experimental crosses are raising as particularly useful resources for QTL mapping, since they allow to address a broader genetic variability and offer the chance of evaluating multiple haplotypes within a single mapping experiment, potentially providing at the same time higher genetic map resolution (Cavanagh *et al*., 2008).

In this work we have described the establishment and the initial utilization of the first multiparental population in durum wheat, named NCCR, which was obtained by experimental crossing four different elite durum wheat cultivars. The final aim of this research was to carry out a multi-trait quantitative trait loci analysis regarding yield and yield-relevant traits in the context of the cultivated background of durum wheat. The population's founders were elite lines which were chosen not only to represent the cultivated genetic diversity, but also because they contributed genetic diversity for important agronomic traits. Neodur is a cultivar characterized by a high number of spikelets per ear, Claudio shows stable high test weight and high number of fertile tillers and it is also characterized by resistance to powdery mildew, Colosseo presents high-yielding ears (in terms of thousands kernel weight) and brown rust resistance, Rascon/Tarro is known for the high number of grains per spikelet. The NCCR population has therefore the potential to enable detecting QTL and estimating QTL allele/haplotype effect directly in a cultivated elite genetic background, providing direct insights about the practical potential usefulness of the identified QTL haplotypes in breeding programs.

In principle, QTL alleles/haplotypes are also intrinsically tested in a more diverse genetic background (up to four functional genetic haplotypes at any locus) than usually obtained in bi-parental crosses, providing a further test to their applicability in breeding.

NCCR-based genetic map and the NCCR population structure

NCCR genetic map spanned 2,663 cM and can be considered as significantly longer than the reference bread wheat consensus maps based on bi-parental populations (Somers *et al*., 2004; Sourdille *et al.,* 2004; Wang *et al.*, 2014). Genetic maps estimated from multiparental populations are likely to show a length expansion when compared to bi-parental maps estimated from a similar number of lines, due to additional opportunities for informative meiosis (Cavanagh *et al*., 2008). Overall, the quality of the NCCR map appeared satisfactory as a monotonic increase was observed when compared to the 90k-bread wheat map (Wang *et al.,* 2014) and an extremely high conservation of marker order have been identified. However, as previously observed (Huang *et al*., 2012) we confirm that building a linkage map in a

multiparental cross has drawbacks. For instance, the genealogical relationships between the four population founders (ie. the presence of IBD chromosome regions between the NCCR founders) inherently prevented to perfectly attribute founders haplotype and crossing-over positions to each population RIL. In our population, as in other multiparental crosses (Huang et al. 2012) this caused an apparently lower-than-expected observed number of recombination events. More precisely, simulations conducted on hexaploid bread wheat and based on a consensus map of total length 2,500 cM (Huang *et al.,* 2012) highlighted that a mean of 51 and 77 recombination events per bi-parental RIL population or four-parent RIL should be expected, respectively. Such expected values, adjusted for the 14 durum chromosomes, approximate to 34 and 51.3 genome-wide recombination events. In NCCR, the genome-wide average number of recombination events per line was 31.5. Similarly to what was found by Huang in the bread wheat multiparental population (Huang *et al*., 2012), the detected recombination events in the NCCR population appeared slightly lower than those expected with a bi-parental RIL. As hypothesized by Huang *et al.* (2012), although the present result seems counterintuitive given the increased NCCR map length, it could be due to the computational difficulty in estimating recombination events in multi-way experimental populations. Moreover, the gaps present in the NCCR map have likely prevented us from detecting all the recombination events present in each RIL. As a matter of fact, the ability to detect recombination events in multiparental crosses relies on the success in estimating the allele founder probabilities along the genome of single RIL. Such estimates are therefore directly and negatively correlated to the percentage of unassigned genomic regions (for instance in long chromosome stretches of IBD), and appear heavily influenced by map gaps. In NCCR map, the percentage of regions not assigned to any founder varied from 9.0% (chr 5A) to 46.1% (chr 3B) per chromosome. On average, for each line, 20.1% of the genome could not be assigned at the chosen threshold probability (0.7) to any founder. In conclusions, these observations seem to indicate that while multiparental crosses are potentially very informative for QTL analysis, an excessively close genealogical relationship between two or more parental lines may seriously hinders the process of map construction and QTL mapping due to the inherent negative effect of such close genealogy on to the algorithms currently in used for linkage map reconstruction and QTL mapping. The inspection of the average percentages of founder assignment for each chromosome highlighted important deviations from the expected theoretical 25%. In particular Claudio's haplotype significantly dropped in frequency on chr 5A and 6B. They therefore might indicate a segregation distortion relative to the Claudio haplotype which should be further investigated.

IBS vs IBD mapping approaches for mapping QTL in NCCR

For a subset of traits (i.e. HD, MD, PH and GY), a comparative QTL analysis have been conducted both with IBD-CIM and IBS-SMA, providing a means for comparing the efficiency of the two mapping methods in effectively modelling and describing the haplotype variation at genomic loci associated to target traits. Indeed, although an algorithm aimed at reconstructing haplotype founder probabilities along the chromosome is essential when constructing a linkage map with multiparental populations, not necessarily QTL mapping on MPPs have to rely on IBD information.

The IBS-SMA analysis we have performed with Tassel software relied on independent F-test at each bi-allelic marker position and can be considered as an IBS mapping approach considering allelic states and regardless of a common-ancestor origin. The main drawback with IBS mapping is that plants with identical marker alleles are grouped together at test SNP positions, irrespective of a possible incomplete linkage disequilibrium between markers and functional QTL alleles (Würschum, 2012). The composite interval mapping which has been performed with mpMap software (Huang et al. 2011) on founder haplotype probabilities, relied on IBD estimates (as it attempted to trace the parental origin of SNP alleles). Such estimates are computed along the chromosomes via a multipoint probability algorithm which does not consider the possible existence of IBS stretches among founders due to common descent. The information carried by founders is therefore modelled as un-related although such assumption is unrealistic with breeding populations, especially if derived from elite materials. Such computation approach would lose possible IBS information among founders, although gaining IBD information as it likely enables to detect QTL also when two founders with local IBS stretches have different effects on the phenotype. At these QTL, a pure IBS-SMA analysis would be misleading as two parents may apparently share the same haplotype but not necessarily the same QTL genotypes. As an example, we have underlined the case of the *QHd.ubo*-2B where IBD-CIM analysis enabled us to discriminate three haplotypes with contrasted functional effects, while SNP-based haplotypes in the target region were only two. The IBD-approach should likely work better in cases where variation at target quantitative loci is multiallelic. As pointed out by Huang et al. (Huang et al 2010) referring to the connected multiparental population, such approach could also better detect variation at 'bi-allelic' quantitative loci determined from clusters of tightly linked QTL. It is noteworthy that in this work the two QTL mapping methods generally agreed in terms of QTL number, map position, effect and proportion of explained variance (both at the single QTL level and as full-model) As obvious, only the IBD-CIM methods enabled us to identify more than two functional haplotypes, if present.

Analysis of QTL results

QTL for grain yield

QTL analysis for grain yield on the combined data over the four environments identified only two QTL *QGy.ubo-2B* and *QGy.ubo-7A*, which were supported by both QTL statistical mapping methods (IBD-CIM and IBS-SMA). In addition, from one to four yield QTL were identified in each single environment. In three cases (at chr 2B for PR11, and at Chr 7A for Cad11 and Arg12) yield QTL from single environments overlapped with the two QTL *QGy.ubo-2B* and *QGy.ubo-7A* with general concordance of direction of allelic effect (in all these cases, the Colosseo QTL allele was characterized by a statistically significant negative effect on yield). It is noteworthy that an important portion of QTL detected for grain yield was found as environment-specific and with relatively minor effect. Accordingly, the full CIM modelling of grain yield in mpMap explained only the 12% of the total genetic variation, while the GY QTL detected by Tassel SMA on chr 2B and 7A showed an R^2 of 5.3 and 7.2%, respectively. The relatively low effect and percentage of phenotypic variance for grain yield QTL in NCCR are likely due to the choice of the four founders, all belonging to the elite cultivated germplasm and generally adapted (although with important differences) to high fertility environments, where the experiments took place. As pointed by Maccaferri *et al.* (2008) QTL for yield and yield-related traits in cereals frequently are found to explain at maximum 10% of the total phenotypic variation. This is particularly true when evaluating segregating materials obtained from elite accessions (Quarrie *et al.,* 2005; Dilbirligi *et al*., 2006). The incomplete correspondence of GY QTL in the same regions of yield-component QTL could have been the result of a fine counterweight balancing among the single components, and to a limited statistical power to detect grain yield QTL of small effect. This notwithstanding, the analysis of yield-component data resulted very informative and contributed to tag several regions carrying QTL alleles of potential importance in durum wheat breeding as discussed below.

QTL for yield components

An extremely interesting QTL was identified for GNE/GNS (grains or kernels per ear and grains per spikelet. *QGns.ubo2A* - *QGne.ubo2A)* on chr 2A. Grain per ear has been identified as probably the most important grain yield component across cereals (Sakamoto and Matsuoka, 2008; Reynolds *et al*., 2009). The *QGns.ubo2A* - *QGne.ubo2A* identified here explained a large portion of phenotypic variation (R^2 = 39.2 and 24.3% for GNS and GNE, respectively), with a clear functional QTL allele increasing the traits value provided by Rascon/Tarro. The same QTL has an important impact on GWT (*QGwt.ubo*-2A, R^2 = 35.7) in this case with the Rascon/Tarro allele strongly decreasing the trait value. At this region, GNE appears modulated in terms of three

functional haplotypes (Neodur – Colosseo *vs* Claudio *vs* Rascon/Tarro) while at the SNE QTL only two functional haplotypes were estimated (Neodur – Claudio – Colosseo *vs* Rascon/Tarro), this supporting the presence of an allelic series at this QTL. It is noteworthy that Rascon/Tarro was previously known for potentially providing favourable alleles for yield potential (i.e. high number of grains per spikelet). The grain yield component GNE, potentially the product of SNE × GNS, appears here completely driven by GNS. However, the understanding of the developmental and physiological mechanisms (higher number of differentiated florets per spike, higher intrinsic floret fertility, etc) behind this QTL requires additional investigations.

Overlaps between NCCR grain yield QTL and yield QTL reported in literature

Grain yield is by far the most investigated trait in terms of QTL analysis in wheat. More than 400 hits were obtained in Web of Science (Thomson Reuters) database after searching using "grain yield QTL wheat" as keywords (on March 27, 2014). Approximately 60% of these manuscripts reported original QTL experimental data. In addition, one manuscript reported the results of QTL meta analysis (Goffinet and Gerber, 2000) for grain yield, ie. an effort aimed at summarizing and synthesizing the huge number of published QTL for this important trait. In this meta analysis, 59 independently published QTL experiments spanning. the 2000-2010 decade were used to collect information on 541 QTL (Zhang *et al*. 2010). This enabled to identify 55 metaQTL, of which 12 metaQTL were considered highly significant. Chr 2 harbored the highest number (21%) of QTL. By comparison with the yield QTL map positions identified in this thesis, none of the grain yield QTL identified across environments corresponded to any major significant yield metaQTL. However, *QGy.ubo.2B* and *QGy.ubo7A* overlapped with three minor yield meta QTL (*MQTL12- MQTL13* and *MQTL49*. Table 3 in Zhang *et al*. 2010). For grain yield QTL identified in single environments, the 2A QTL identified in Cad12 and Arg12 coincided with *MQTL10* (Table 3 in Zhang *et al*. 2010) at approx position 90-100 cM. It is noteworthy that the major QTL for grain number per spikelet and per ear (*QGns.ubo2A* - *QGne.ubo2A*) identified in this thesis also co-maps at this region. We therefore further checked the single QTL experiments utilized for meta analysis at this chr 2AS region in order to verify how often QTL for grain number per ear underlined the *MQTL10* for grain yield. Indeed, *MQTL10* resulted from QTL identified in four independent studies (Huang *et al* 2003; Kumar *et al* 2007; Li *et al*. 2007; Wang *et al.* 2009) and in all such studies a QTL for grain number per ear co-mapped with *MQTL10,* Zhang *et al*. (2010).

It appears therefore that our major *QGns.ubo2A* - *QGne.ubo2A* QTL on 2AS co-maps with several relevant QTL for the same yield component and for yield identified in independent genetic backgrounds (all different bread wheat experimental crosses). This suggests that further studies (eg. fine mapping and cloning) on the *QGns.ubo2A* - *QGne.ubo2A* QTL could be of general

potential interest for wheat genetic improvement. From 2010 to March 2014, approximately 40 QTL mapping experiments in addition to the ones summarized in Zhang *et al* (2010) have been published. Although it goes beyond the objective of this thesis to formally compare (eg. by means of meta analysis) our QTL with all such QTL, the following interesting QTL overlaps were observed. Our *QGy.ubo.2B yi*eld QTL overlapped with several QTL and QTL clusters for yield and phenology, likely due to the segregation of major loci influencing phenology (eg. *Ppd*) on 2B, as already discussed. Of more interest, Bennet *et al.* (2012) and Wu *et al*. (2012) reported grain yield or grain yield component QTL on chr 7A which overlap with our grain yield QTL *QGy.ubo7A*.

QTL for phenology-related traits

Genes controlling phenology are key for wheat breeding as they are among the main factors determining cultivar adaptation to a given environment and are highly correlated with grain yield. In NCCR, phenology-related traits (HD and MD, combined value across environments) showed a relatively low correlation value (r = 0.26) with grain yield, despite the NCCR population presented a quite extended range of variation for HD (min value = 132.3 days from sowing, max value = 149.0 days from sowing). Slightly higher correlation between HD and yield components were observed, in particular regarding SNE (r = 0.46) and GWE (r = 0.40).

The main QTL affecting both HD and MD on chr 2A, 2B and 7A were likely positioned on the well-known *PpdA1*-2A, *PpdB1*-2B and *Vrn3*-7A loci affecting photoperiod sensitivity and vernalization requirements. Evidences for this were provided by an analysis of co-location of the QTL supporting intervals in the NCCR map and in previously reported QTL and meta-QTL attributed to *Ppd* and *FT* loci in hexaploid and tetraploid wheat (Griffiths *et al*., 2009). This was carried out based on the unique feature of the iSelect wheat 90k assay of being derived from the transcribed portion of the genome.

The *QHd.ubo-2A* corresponded to the meta-QTL 1 for HD reported by Griffiths *et al.* (2009) in chr 2A. Similarly, in the durum wheat Kofa x Svevo, QHd-2A.2 (Maccaferri *et al*., 2008) was mapped to this position. *QHd.ubo-2B* corresponded well with *Ppd-B1* mapped by Bennett *et al*. (2012). In NCCR, *QHd.ubo-7A* was mapped to a relative interval compatible to those reported in Bennett *et al*. (2012), in Hanocq *et al.* (2007), and in Griffitths *et al.* (2009), with 7A meta-QTL1 located to an interval of 9-19 and 7A meta-QTL2 to 33.5-49.03.

Among the QTL that affected phenology, the only QTL that also showed strong related effects on PH was *QHd.ubo-7A,* for which the direction of the effects on PH were consistent with the observed effects on HD (i.e. the earliest founder effect associated to Colosseo was associated to a

shortest height of 3.91 cM). Excluding *QHd.ubo-7A*, *QPh.ubo-2B* was the strongest QTL for PH *per* se, with R^2 = 7.7% across environments and appreciable effects on MD (significant in the IBS-SMA, R^2 of 4.1%) and particularly GY (R^2 of 7.0%) across environments. *QPh.ubo-2B* is therefore to be considered a major QTL for plant vigour independent to the regulation of heading. It is located ca. 50 cM proximal from the location of *Ppd-B1*. For this QTL, the plant height effects were positively correlated with yield effects.

QTL for protein content

Four QTL involved in the expression of GPC have been identified, all of them associated with significant QTL for grain yield (chr 2B, chr 7A. In the latter case, the location of the GY peak remains approximately defined) or grain yield components (chr 2A, 4B). Generally, an opposed pattern in terms of allelic effect is detected for GPC and GY or GY-components (Zanetti *et al*., 2001; Blanco *et al*., 2002, 2006). A negative correlation between productivity and GPC is found in most segregating populations likely due to pleiotropic genetic effects and complex interdependent development patterns (Blanco *et al*., 2011). Modern high-yielding cultivars are characterized by a relative low grain protein content as practical breeding programs indirectly selected for decreasing GPC, while aiming to enhance grain yield. Thus, identifying QTL that appear to influence GPC independently from variation in yield components is particularly promising from an applied breeding perspective. Our QTL cluster on chr 2A provides an interesting case. At this region, the GPC and GWT QTL unexpectedly showed the same direction of genetic effect (high GPC corresponded to high GWT). However, the concurrent clustering of QTL for yield-components related to ear morphology clarified that both GPC and GWT were likely driven by the segregation of the major QTL for GNS/GNE (*QGns.ubo-2A/QGne.ubo-2A*). Thus, at this region, a higher number of kernels per ear was associated with lower GPC and GWT, as commonly observed.

QTL clusters

Multi-trait QTL analysis enabled us to identify and highlight QTL clusters that showed concomitant effect on several agronomical important traits. Based on the IBD-CIM QTL analysis, which provides the estimate of allele/haplotype effects for all four founders across the genome, we have additionally investigated such clusters for coherence in terms of patterns of allelic/haplotype effects. One typical example is provided by heavy influence of phenology (and therefore phenology-related QTL) on crop yield in cereals (Reynolds and Tuberosa, 2008). The analysis of QTL-clusters should in principle enable us to characterize the adaptive values (in terms of effects on yield) of QTL for heading and maturity, and at least in principle provides the opportunity to identify QTL for yield or yield components which were relatively independent

from phenology. These QTL would be extremely important for breeding purposes (Reynold *et al.*, 2009; Pinto *et al.,* 2010). Main QTL clusters where identified on chr 2A, 2B, 4B and 7A.

The cluster of QTL on chr 2A distal involved major loci associated to phenology-related traits (HD, R^2 = 22.0%; MD, R^2 = 11.5% with IBD-CIM and HD, R^2 = 22.9%; MD, R^2 = 10.5% with IBS-SMA) and QTL for FLER (R^2 =3.0%), FLG (R^2 =3.8%), NDVI_57 (R^2 =6.9%) and GWE (R^2 =2.4%).

The QTL cluster detected on chr 2A proximal (a region associated to MD with IBS-SMA combined data analysis and to MD and GY in Cad12, GY in Arg12) appeared as extremely enriched by a multi-trait QTL analysis including morpho-physiological, yield-components and quality data. Apart from detecting QTL for EGC (R^2 = 8.8%), NDVI_31 (R^2 = 16.6%), NDVI_40 (R^2 = 8.2%), in the same cluster we could also map yield-component QTL which appeared as extremely appealing. First of all, we noticed a genomic region highly associated with GNS (*QGns.ubo*-2A, R^2 = 39.2%), SNE (*QSne.ubo*-2A, R^2 = 16.1%), GNE (*QGne.ubo*-2A, R^2 = 24.3%) and GWT (*QGwt.ubo*-2A, R^2 = 35.7%). A relatively milder association have been described for FTsm (R^2 = 3.9%). Interestingly, such QTL cluster appeared to be driven by the favourable effect of the haplotype provided in the four-way cross by Rascon/Tarro, which is a cultivar characterized by favourable alleles for yield potential (i.e. high number of grains per spikelet). It is very interesting to note that the two single-environment GY QTL detected in the same region (Cad12 and Arg12) showed contrasting genetic effects in terms of haplotypes. These could perhaps contribute to explain the lack in GY QTL which was observed when analysing combined data.

Regarding the cluster detected on chr 2B proximal, heading date seemed not to influence *QGY.ubo*-2B, which co-mapped with a relative high-effect QTL for plant height. The peak of *QGY.ubo*-2B resulted 51.9 cM apart from *QHd.ubo-2B* and a very mild potential MD effect on *QGY.ubo*-2B was detected by IBS-SMA only (in terms of *QMd.ubo-2B.2*). This QTL cluster showed significant effects mainly for PH on three environments (Cad11, Cad12 and Pr11, with R^2 equal to 9.3, 3.7, 5.7%, respectively) and for GY in Pr11 (R^2 = 3.4%). Conversely, the QTL for days to maturity detected by IBS-SMA has not been revealed by any single-environment CIM analysis at the target threshold.

On chr 4B we observed two QTL clusters. In the distal one, a QTL for SNE was mapped (R^2 = 9.9%) as unique yield component in a region characterized by HD and MD QTL, as well as by early-vigour QTL (EGC, R^2 = 4.3% and FLER, R^2 =5.6%). The proximal 4B QTL cluster involved various yield-related QTL (for FTsm, R^2 =6.0%; GNS, R^2 = 1.7%; GNE, R^2 =3.0%, GWE, R^2 = 3.7% and GWT, R^2 = 3.5%), a GPC QTL (R^2 = 5.6%). Analysing data from Arg12, we found a GY QTL (R^2 =6.0%) mapping in the same region. The directions of the genetic effects appeared conserved and coherent with respect to a QTL for MD (detected both by IBD-CIM and IBS-SMA with R^2 =

3.9% with both methods). Therefore, although GNS and SNE seem here to be determined by different genetic factors, the variation in GNE could be hypothesized as influenced by the MD locus. The confidence intervals of the MD and yield-related QTL were distributed over a large chromosome portion and this prevented us from drawing final conclusions about the examined cluster.

On chromosome 7A, *QHd.ubo-7A* and *QMd.ubo-7A* appeared to influence both PH and GY, a trend confirmed as well in the QTL analysis performed on single environments. The cluster has been additionally populated by QTL for yield-components (SNE, R^2 = 7.2%; GNE, R^2 = 9.2%; GNS, R^2 =4.2%, GWE, R^2 =8.6%). At this cluster, the *post-hoc* test on the allelic effects, indicated that three functional alleles are likely to be present (Neodur - Claudio *vs* Colosseo *vs* Rascon/Tarro). As a greater number of functional haplotypes are segregating, the pattern of variation at this region seems less easy to be interpreted than that involved in chr 2A proximal cluster. Anyway, SNE trait variation appears to have a major role in the pattern of allelic effects present at this cluster, and was likely correlated to the pattern of effects at the HD and MD loci. The most likely explanation is that the Colosseo haplotype as an early-haplotype in terms of development-related QTL, determines a relatively major decrease both in SNE and GNS. At the target region, the Colosseo haplotype effect appears more negative the Rascon/Tarro one, while the estimated Neodur – Claudio common haplotype apparently increased GNS.

Possible impact of NCCR analysis on improving yield potential

In this work, we have had the opportunity to analyse the effect of the segregation of haplotypes deriving from parents characterized by very different features in terms of yield potential, particularly related to the sink-source balance. Among NCCR founders, Neodur (late genotype) was characterized by a high number of spikelet per ear, while Rascon (early genotype) is known as a cultivar producing a high number of grains per spikelet. This has enabled us to map QTL for ear-morphology traits (the most important detected loci were mapped on chr 2A, 4B, 7A), to discriminate the relative contribution of SNE and GNS on GNE, and to describe their interconnection in terms of four founder haplotypes. The QTL for GNS detected on chr 2A appeared particularly promising as likely driven by a favourable effect of a relatively rare (in terms of the sampled founder) haplotype, which was present at a ratio 1 *vs* 3 (i.e. Rascon/Tarro vs Neodur, Claudio and Colosseo). The detected allelic effect for a yield-potential QTL was estimated in elite genetic background, thus in a framework close to real breeding practices. In order to describe the physiological factors determining GY genetic gains in crops, yield potential has been alternatively described in terms of numerical components or harvest index. Since the Green revolution, genetic improvement in wheat has been mainly achieved as a consequence of

an indirect genetic shift towards greater harvest index (Fischer, 1984) likely caused by an increase of grain number per unit area. On the other hand, it has been reported (Loss *et al*., 1989; Slafer and Andrade, 1989) that individual grain weight has even been reduced by wheat breeding during the past century without final yield being seriously affected (Pfeiffer et al., 2000). The importance of such trade-off between the two main yield components have led wheat scientists to look for alternative ways to increase crop yield potential. In particular, several physiological evidences indicate that grain yield in wheat is sink-limited (Miralles *et al*., 2000; Miralles and Slafer 2007; Reynolds *et al*., 2007) and that a promising path for raising yield potential relies in improving the source/sink balance. A larger genetic plasticity both in terms of kernel size and number of kernel per ear could likely contribute to efficiently accommodate an increased carbon remobilization from source tissues. There is therefore a growing interest in better understanding the role of spike morphological traits in determining grain yield genetic gains, since a wheat ideotype should present long spikes with high spikelet number per ear and high number of grains per spikelet (Gaju *et al*., 2009). In particular, grain number per ear (which in this work has been denoted as GNE) it is a complex trait mainly determined by two individual components: spikelet numb per ear (SNE) and Grain number per spikelet (GNS) whose genetic control has not yet been well characterized. It has been noticed that these two components generally appear negatively correlated and that early genotypes generally produce a lower number of spikelet per ear (Pugsley, 1966; Worland and Law, 1986; Sourdille *et al*., 2000). The developmental basis of this phenomenon is unknown. For instance it is not clear the relative contribution of floret differentiation and floret abortion to the observed spike fertility in different wheat genotypes. Experimental crosses aimed at inducing higher grain sink capacity have been performed utilizing genetic sources characterized by multi-ovary florets (Skovmand *et al*., 2000) and branched spikelets (Dencic, 1994). In particular, Dencic (1994) succeeded in significantly increasing ear length, number of spikelets per ear, grains per spikelet, and number of grains per ear while he highlighted a negative correlation between number of grains per ear and individual grain weight. In our case, Rascon/Tarro has been chosen as founder of the multiparental population as it is known to be characterized by an high yield potential in terms of number of grain per spikelet.

CONCLUSIONS

In this study we have carried out a multi-trait multi-environment QTL analysis deploying the NCCR population, a new multiparental population in durum wheat including 338 RILs. NCCR was developed starting from crossing four cultivars, Neodur, Claudio, Colosseo and Rascon/Tarro, which represent the cultivated elite germplasm and are characterized by different traits of agronomic value. QTL mapping experiments on this population should provide insight into the genetic control of yield in elite durum materials. At the same time, effect of relevant QTL alleles are discovered and estimated in a context closer to real plant breeding.

NCCR was characterized in four-environment field trials for 18 traits, including yield, yield-components, morpho-physiological and quality traits. High-throughput genotyping was obtained based on a recently-developed wheat-dedicated 90k SNP chip in which we have specifically contributed a "cluster file" algorithm required for correct genotyping of tetraploid wheats. Such durum wheat-tailored algorithm could serve the entire durum wheat community in future genotyping experiment involving the same chip.

The NCCR-based linkage map (2,663 cM) covered all the 14 wheat chromosomes and has been quality-checked with a recently released bread wheat consensus map. Although the quality appeared highly satisfactory, intervals with relatively low marker density remained. Regions of Identity-by-descent (IBD) among the four founders are probably the main cause of the occasional difficulties in linkage map construction as well as of the relatively high rate (approx. 20%) of unassigned genotype along RIL chromosomes.

Among the many QTL identified, we have mapped two QTL for grain yield across environments and 23 QTL for grain yield components. A novel major QTL for number of grain per spikelet/ear was mapped on chr 2A and shown to control up to 39% of phenotypic variance in this cross. More generally, our QTL results confirmed the importance of relatively minor and environment-specific QTL in controlling grain yield in the elite durum wheat germplasm. In addition, the analysis of the main QTL cluster offered by the multi-trait QTL analysis confirmed the strong influence of phenology-related loci on yield, although QTL for yield and yield components which were relatively independent from phenology were also identified. For all detected QTL, it was possible to estimate the number of functionally different QTL haplotypes, which was on average 2.3 per locus, providing an estimate to the residual useful genetic variation still present in the durum wheat elite germplasm.

As a future perspective, the major QTL for number of grain per spikelet/ear identified in this study is worth of additional investigation. It should be confirmed (e.g. by verification in diverse association panels), further functionally characterized (e.g. by testing its effect on yield after isogenization in different genetic backgrounds) and eventually cloned. Results based on this

study also indicated that linkage and QTL analysis based on multiparental population can still be improved by developing better algorithms improving founders haplotype reconstruction in the progeny. On the other hand, our results indicate that the IBD relationships among founders should be carefully considered when planning such highly demanding (in terms of time and funding) multiparental resources.

REFERENCES

Akhunov E, Nicolet C, Dvorak J, 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theoretical and Applied Genetics, 119: 507-517.

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Hao H, Huo N, Lazo GR, Luo M-C, Ma YQ, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, and Dvorak J. 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics, 11: 702.

Araus JL, Slafer GA, Reynolds MP, Royo C, 2002. Plant breeding and drought in C3 cereals: What should we breed for? Annals of Botany, 89: 925-940.

Arumuganathan K, Earle ED, 1991. Nuclear DNA content of some important plant species. Plant Molecular Biology Reports, 9: 208-219.

Austin RB, Bingham J, Blackwell RD, Evans LT, Ford MA, Morgan CL, Taylor M, 1980. Genetic improvement in winter wheat yield since 1900 and associated physiological changes. J. Agric. Sci. 94:675–689.

Balding DJ, 2006. A tutorial on statistical methods for population association studies. Nature Genetics, 7: 781-791.

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Redoña E, Leung H, 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice, 6: 11.

Bao JS, Wu YR, Hu B, Wu P, Cui HR, Shu QY, 2002. QTL for rice grain quality based on a DH population derived from parents with similar apparent amylose content. Euphytica, 128: 317-324.

Bardol N, Ventelon M, Mangin B, Jasson S, Loywick V, Couton F, Derue C, Blanchard P, Charcosset A, Moreau L, 2013. Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (Zea mays L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. Theoretical and Applied Genetics, 126: 2717-2736.

Barker GLA, Edwards KJ, 2009. A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. Plant Biotechnol J, 7: 318-325.

Barrett JC, Fry B, Mallerand J, Daly MJ, 2005. Haploview:analysis and visualization of LD and haplotype maps. Bioinformatics, 21:263-265.

Beavis WD, 1994. The power and deceit of QTL mapping experiments: lessons from comparative QTL studies. Proceedings of the Annual Corn and Sorghum Industry Research Conference, 49: 250-265.

Bennett D, Izanloo A, Reynolds M, Kuchel H, Langridge P, Schnurbusch T, 2012. Genetic dissection of grain yield and physical grain quality in bread wheat (*Triticum aestium*L.) under water-limited environments. Theoretical and Applied Genetics, 125: 255–271.

Bentsink L, Hansona J, Hanhart CJ, Blankestijn-de Vries H, Coltrane C, Keizer P, El-Lithy M, Alonso-Blanco C, de Andrés MT, Reymond M, van Eeuwijk F, Smeekens S, Koornneef M, 2010. Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. Proceedings of the Natural Academy of Sciences, 107: 4264-4269.

Bernardo R, Moreau L, Charcosset A, 2006. Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection. Crop Science, 46: 1972-1980.

Bernardo R, 2008. Molecular marker and selection for complex traits in plants: learning from the last 20 years. Crop Science, 48: 1649-1664.

Blanco A, Giorgi B, Perrino R, Simeone R, 1990. Risorse genetiche e miglioramento della qualità del frumento duro. Agricoltura Ricerca, 114: 41-58.

Blanco A, Pasqualone A, Troccoli A, Di Fonzo N, Simeone R, 2002. Detection of grain protein content QTLs across environments in tetraploid wheats. Plant Molecular Biology, 48: 615-623.

Blanco A, Simeone R, Gadaleta A, 2006. Detection of QTLs for grain protein content in durum wheat. Theoretical and Applied Genetics, 112: 1195-1204.

Blanco A, Mangini G, Giancaspro A, Giove S, Colasuonno P, Simeone R, Signorile A, De Vita P, Mastrangelo A, Cattivelli L, Gadaleta A, 2011. 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. Molecular Breeding, 30: 79-92.

Bonnin I, Rousset M, Madur D, Sourdille P, Dupuits C, Brunel D, Goldringer I, 2008. FT genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. Theoretical and Applied Genetics, 116: 383-394.

Borras L, Slafer GA, Otegui ME, 2004. Seed dry weight response to source-sink manipulations in wheat, maize and soybean: a quantitative reappraisal. Field CropsResearch 86: 131-146.

Bozzini A, 1988. Origin, distribution, and production of durum wheat in the world. In: Fabriani C, Lintas C (Eds.), Durum Wheat: Chemistry and Technology. Saint Paul. Minesota. USA.

Bradbury P, Parker T, Hamblin MT, Jannink JL, 2011. Assessment of power and false discovery rate in genome-wide association studies using the Barley CAP germplasm. Crop Science, 51: 52–59.

Bradshaw HD, 1996. Molecular genetics of populus. In: Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM (Eds.), Biology of populus and its implications for management and conservation, Part 1. Ottawa, NRC Research Press, 183-199.

Breseghello F, Sorrels ME, 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science, 46: 1323-1330.

Broman KW, 2001. Review of statistical methods for QTL mapping in experimental crosses. Laboratory Animals, 30: 44-52.

Broman KW, Wu H, Sen S, Churchill GA, 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19: 889-890.

Broman KW, 2005. The genomes of recombinant inbred lines. Genetics, 169: 1133-1146.

Broman KW, Sen S, 2009. A guide to QTL mapping with R/qtl. Springer, New York.

Brookes AJ, 1999. The essence of SNPs. Gene 234: 177-186.

Brooking IR, 1996. Temperature response of vernalization in wheat: a developmental analysis. Annals of Botany, 78: 507-512.

Browning BL, 2008. PRESTO: Rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies. BMC Bioinformatics, 9: 309.

Browning BR, Browning BL, 2011. Haplotype phasing: existing methods and new developments. Nat. Rev. Genet., 12: 703–714.

Buckler ES, Thornsberry JM, 2002. Plant molecular diversity and applications to genomics. Current Opinions in Plant Biology, 5: 107-111.

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, 2007. Analysis of mixed models for S language environments, ASReml-R reference manual Release 2. QLD Department of Primary Industries and Fisheries, Brisbane, QLD. Training and Development Series, No QE02001.

Calderini DF, Dreccer MF, Slafer GA, 1995. Genetic improvement in wheat yield and associated traits. A re-examination of previous results and the latest trends. Plant Breeding, 114: 108–112.

Calus MPL, Veerkamp RF, 2007. Accuracy of breeding values when using and ignoring the polygenic effects in genomic breeding value estimation with a marker density of one SNP per cM. J. Anim. Breed. Genet., 124: 362-368.

Cavanagh C, Morell M, Mackay I and Powell W, 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr. Opin. Plant Biol., 11: 215-21.

Cavanagh C, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, Lopez da Silva M, Bockelman, Luther Talbert, James A. Anderson, Susanne Dreisigacker, Stephen Baenziger H, Carter A, Korzun V, Laurent Morrell P, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E, 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci., 110: 8057–8062.

Chakravarti A, Lasher LK, and Reefer JE, 1991. A maximum likelihood method for estimating genome length using genetic linkage data. Genetics, 128:175–182.

Cheverud JM, 2001. A simple correction for multiple comparisons in interval mapping genome scans. Heredity, 87: 52–58.

Chouard P, 1960. Vernalization and its relations to dormancy. Annual Review of Plant Physiology, 11: 191-238.

Churchill GA, Doerge RW, 1994. Empirical threshold values for quantitative trait mapping. Genetics, 138: 963-971.

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, *et al.,* 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet, 36: 1133–1137.

Clark AG, 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol., 7: 111-122.

Cochran WG, Cox GM, 1960. Experimental designs. John Wiley and Sons, New York, USA.

Comadran J, Ramsay L, MacKenzie K, Hayes P, Close TJ, Muehlbauer G, Stein N, Waugh R, 2011. Patterns of polymorphism and linkage disequilibrium in cultivated barley. Theoretical and Applied Genetics 122:523–531.

Condon AG, Richards RA, Rebetzke GJ, Farquhar GD, 2004. Breeding for high water-use efficiency. Journal of Experimental Botany, 55: 2447-2460.

Crossa J, de los Campos G, Pérez P, Gianola D, Burgueño JL, Araus, D, Makumbi, R.P. Singh, Dreisigacker S, Yan J, Arief V, Banziger M, Braun HJ, 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers, 186: 713-724.

Daetwyler H. D, Pong-Wong R, Villanueva B, Woolliams JA, 2010. The impact of genetic architecture on genome-wide evaluation methods. 185: 1021–1031.

Daly MJ, Rioux, J, Schaffner, S, and Hudson T, 2001. High resolution haplotype structure in the human genome. Nat.Genet., 29: 229–232.

Darvasi A, Soller M, 1997. A simple method to calculate resolving power and confidence interval of QTL map location. Behavior Genetics, 27: 125-132.

Dencic S, 1994. Designing a wheat ideotype with increased sink capacity. Plant Breeding, 112: 311-317.

Dilbirligi M, Erayman M, Campbell BT, Randhawa HS, Baezingeret PS, Dweikat I, Gill KS, 2006. High-density mapping and comparative analysis of agronomically important traits on wheat chromosome 3A. Genomics, 88: 74-87.

Distelfeld A, Tranquilli G, Li C, Yan L, Dubcovsky J, 2009. Genetic and molecular characterization of the VRN2 loci in tetraploid wheat. Plant Physiology, 149: 245-257.

Doerge RW, Zeng Z, Weir BS, 1997. Statistical issues in the search for genes affecting quantitative traits in experimental populations. Statistical Science, 12: 195-219.

Doerge RW, 2002. Multifactorial genetics: Mapping and analysis of quantitative trait loci in experimental populations. Nature Review Genetics, 3: 43-52.

Dubcovsky J, Loukoianov A, Fu D, Valarik M, Sanchez A, Yan L, 2006. Effect of photoperiod on the regulation of wheat vernalization genes VRN1 and VRN2. Plant Mol. Biol. 60, 469-480.

Dubcovsky J, Dvorak J, 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. Science, 316: 1862-1866.

Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC, 2006. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexa-ploid wheat. Mol Biol Evol, 23: 1386-1396.

East EM, 1916. Studies on size inheritance in nicotiana. Genetics, 2: 164-176.

Edwards D, Batley J, Cogan NOI, Forster JW, Chagné D, 2007. In: Oraguzie NC, et al.(Eds.), Single Nucleotide Polymorphism discovery, Association mapping in plants. Springer, New York, 53-76.

Efron B, 1979. Bootstrap methods: another look at the jack-knife. Annals of Statistics, 7: 1-26.

Evans LT, Wardlaw IF, Fischer RA, 1975. Wheat. In: Evans LT (Ed.), Crop Physiology. Cambridge, UK, Cambridge University Press, 101-149.

Evenson RE, Gollin D, 2003. Assessing the Impact of the Green Revolution, 1960 to 2000. Science, 300: 758-762.

Ewing B, Hillier L,Wendl MC, Green P, 1998b. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res., 8: 175-185.

Ewing B, Green P, 1998a. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res., 8: 186-194.

Excoffier L, Slatkin M, 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol., 12: 921-927.

Falconer DS, Mackay TFC, 1996. Introduction to quantitative genetics. 4th ed. Harlow, UK, Longman Group.

Fallin D, Schork N, 2000. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am. J. Hum. Genet., 67: 947-959.

Filella I, Serrano L, Serra J, Peñuelas J, 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. Crop Science, 35: 1400-1405.

Fischer RA, 1984. In: Smith, WH, Banta SJ (Eds.), Wheat, Symposium on Potential Productivity of Field Crops Under Different Environments. IRRI: Los Baños, 129-153.

Fischer RA, 2009. Number of kernels in wheat crops and the influence of solar radiation and temperature. The Journal of Agricultural Science, 105: 447.

Flavell RB, Rimpau J, Smith DB, 1977. Repeated sequence DNA relationships in 4 cereal genomes. Chromosoma, 63: 205–222.

Flint J, Valdar W, Shifman S, Mott R, 2005. Strategies for mapping and cloning quantitative trait genes in rodents. Nature Reviews Genetics, 6: 271-286.

Flint-Garcia SA, Thornsberry JM, Buckler ES, 2003. Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54: 357-374.

Flood RG, Halloran GM, 1986. Genetics and physiology of vernalization response in wheat. Advances in Agronomy, 39: 87-124.

Foulkes MJ, Sylvester-Bradley R, Worland AJ, Snape JW, 2004. Effects of a photoperiod-response gene Ppd-D1 on yield potential and drought resistance in UK winter wheat. Euphytica, 135: 63-73.

Frederick JR, Bauer PJ, 1999. Physiological and numerical components of wheat yield. Wheat Ecology and Physiology of Yield Determination, 45-65.

Gabriel SB, et al., 2002. The structure of haplotype blocks in the human genome. Science, 296: 2225–2229.

Gaju O, Reynolds MP, Sparkes DL, Foulkes MJ, 2009. Relationships between large-spike phenotype, grain number and yield potential inspring wheat. Crop Science, 49: 961-973.

Ganal MW, Roder MS, 2007. Microsatellite and SNP markers in wheat breeding, Vol. 2. In: Varshney RK, Tuberosa R, (Eds.), Genomic assisted crop improvement: genomics applications in crops. The Netherlands: Springer, 1-24.

Ganal MW, Altmann T, Roder MS, 2009. SNP identification in crop plants. Current Opinion in Plant Biology, 12: 211–217.

Gao X, Becker LC, Becker DM, Starmer JD, Province MA, 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic Epidemiology, 34: 100–105.

Geldermann H, 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theoretical and Applied Genetics, 46: 319-330.

Girma K, Martin KL, Anderson RH, Arnall DB, Brixey KD, Casillas MA, Chung B, Dobey BC, Kamenidou SK, Kariuki SK, Katsalirou EE, Morris JC, Moss JQ, Rohla CT, Sudbury BJ, Tubana BS, Raun R, 2006. Mid-season prediction of wheat-grain yield potential using plant, soil, and sensor measurements. Journal of Plant Nutrition, 29: 873-897.

Goffinet B, Gerber S, 2000. Quantitative trait loci: a meta-analysis. Genetics, 155: 463-473.

Griffiths S, Simmonds J, Leverington M, Wang Y, Fish L, Sayers L, Alibert L, Orford S, Wingen L, Herry L, Faure S, Laurie D, Bilham L, Snape J, 2009. Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theoretical and Applied Genetics, 119: 383–395.

Groos C, Robert N, Bervas E, Charmet G, 2003. Genetic analysis of grain protein content, grain yield and thousand-kernel weight in bread wheat. Theoretical and Applied Genetics, 106: 1032-1040.

Gupta PK, Rustgi S, Kulwal PL, 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Molecular Biology, 57: 461-485.

Gusfield D, 2001. Inference of haplotypes from samples of diploid populations: Complexity and algorithms. J. Comp. Biol., 8: 305-323.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ, 2011. Extension of the Bayesian alphabet for genomic selection. 12: 186.

Haley CS, Knott SA, 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity, 69: 315-324.

Hanocq E, Laperche A, Jaminon O, Laine AL, Le Gouis J, 2007. Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. Theoretical and Applied Genetics, 114: 569-584.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C,Hochu I, Poirier S, Santoni S, Glémin S, David J, 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol. Biol. Evol., 24: 1506-1517.

Hay RKM, Walker AJ, 1989. An introduction to the physiology of crop yield. Longman Scientific & Technical - Science.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME, 2009. Invited review: Genomic selection in dairy cattle: progress and challengess. J. Dairy Sci., 92: 433-443.

Heffner EL, Sorrells ME, Jannink JL, 2009. Genomic Selection for Crop Improvement. Crop Science, 49: 1-12.

Higgins JA, Bailey PC, Laurie DA, 2010. Comparative genomics of flowering time pathways using Brachypodium distachyon as a model for the temperate grasses. PLoS ONE 5: e10065.

Holland JB, 2007. Genetic architecture of complex traits in plants. Current Opinion in Plant Biology, 10: 156-161.

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A, 1997. More on the efficiency of marker-assisted selection. Theoretical and Applied Genetics, 95: 1181-1189.

Hospital F, Goldringer I, Openshaw S, 2000. Efficient marker-based recurrent selection for multiple quantitative trait loci. Genetics Research, 75: 357–368.

Huang BE, George AW, 2011. R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. Bioinformatics, 27: 727–729.

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell M, Cavanagh C, 2012. A Multiparent Advanced Generation Inter-Cross population for genetic analysis in wheat. Plant Biotechnology Journal, 10: 826-839.

Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M, Eeuwijk FV, 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proceedings of the National Academy of Science, 108: 4488-4493.

Hudson RR, 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol., 7: 1-44.

Isidro J, Alvaro F, Royo C, Miralles DJ, Garrido LF, 2011. Changes in apical development of durum wheat caused by breeding during the 20th century: analysis by phases and its implications for yield formation. Annals of Botany, 107: 1355-1366.

Jannink JL, Bink M, Jansen RC, 2001. Using complex plant pedigrees to map valuable genes. Trends in Plant Science, 6: 337-342.

Jannink JL and Jansen RC, 2001. Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics, 157: 445-454.

Jannink JL and Walsh B, 2002. Association mapping in plant populations. In: Quantitative Genetics, Genomics and Plant Breeding, edited by M. S. Kang. CAB International, New York, 59–68.

Jansen RC, 1993. Interval mapping of multiple quantitative trait loci. Genetics, 135: 205-211.

Jansen RC, Stam P, 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics, 136: 1447-1455.

Jansen RC, Jannink JL, Beavis WD, 2003. Mapping quantitative trait loci in plant breeding populations: Use of parental haplotype sharing. Crop Sci, 43: 829-834.

Jeffreys AJ, Kauppi L, Neumann R, 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat. Genet., 29: 217-222.

Kao C, Zeng Z, Teasdale RD, 1999. Multiple interval mapping for quantitative trait loci. Genetics, 152: 1203-1216.

Kato K, Yamagata H, 1988. Method of evaluation of chilling requirement and narrow-sense earliness of wheat cultivars. Japanese Journal of Breeding, 38: 172-186.

Kato K, Yokoyama H, 1992. Geographical variation in heading characters among wheat landraces, *Triticum aestivum* L., and its implication for their adaptability. Theoretical and Applied Genetics, 84: 259-265.

Kearsey MJ, Farquhar AGL, 1998. QTL analysis; where are we now? Heredity, 80: 137-142.

Kihara H, 1919. Uber cytologische Studien bei einigen Getreidearten. Mit. 1. Spezies- Bastarde des Weizens und Weizenmggen-Bastarde. Tokyo Botanical Magazine, 33: 17-38.

Kihara H, 1924. Cytologische und genetische Studiell bei wichtigen Getreidearten mit besonderer Riicksicht auf das Verhaltell der Chromosomen und die Sterilitat in den Bastarden. Memoirs of the College of Science, Kyoto Imp. Univ., Series B, 1: 1-200.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R, 2009. A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genetics 5: e1000551.

Kruglyak L, 1999. Prospects for whole genome linkage disequilibrium mapping of common disease genes. Nature Genetics, 22: 139-144.

Kuchel H, Williams KJ, Langridge P, Eagles HA, Jefferies SP, 2007. Genetic dissection of grain yield in bread wheat. I. QTL analysis. Theoretical and Applied Genetics, 115: 1029–1041.

Kumar N, Kulwal PL, Balyan HS, Gupta PK, 2007. QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. Molecular Breeding, 19: 163–177.

Kumar S, Sharma V, Chaudhary S, Tyagi A, Mishra P, Priyadarshini A, Singh A, 2012a. Genetics of flowering time in bread wheat Triticum aestivum: complementary interaction between vernalization-insensitive and photoperiod-insensitive mutations imparts very early flowering habit to spring wheat. J Genet., 91: 33-47.

Kumar S, Banks TW, Cloutier S, 2012b. SNP discovery through next-generation sequencing and its applications. Int. J. Plant Genomics BMC Genomics, 13: 684.

Lander ES, Green P, 1987. Construction of multilocus genetic linkage maps in humans. Proceedings of National Academy of Science, 84: 2363-2367.

Lander ES, Botstein D, 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 121: 185-199.

Landserg JJ, 1977. Effects of weather on plant development. In: Landserg J, Cutting CV, (Eds), Environmental Effect on Crop Physiology. Accademic Press, London, 289-307.

Lange K, Boehnke M, 1982. How many polymorphic genes will it take to span the human genome? Am. J. Hum. Genet., 34: 842–845.

Law CN, Worland AJ, 1997. Genetic analysis of some flowering time adaptive traits in wheat. New Phytologist, 137: 19-28.

Le Gouis J, Bordes J, Ravel C, Heumez E, Faure S et al., 2012 Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. Theor. Appl. Genet., 124: 597-611.

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, 2002. Expanding the genetic map of maize with the intermated B73 × Mo17 (IBM) population. Plant Molecular Biology, 48: 453-461.

Leroux D, Rahmani A, Jasson S, Ventelon M, Louis F, Moreau L, Mangin B Clusthaplo, 2014. A plug in for MCQTL to enhance QTL detection using ancestral alleles in multi cross design. Theor Appl Genet, 127: 921-933.

Long JC, Williams RC, Urbanek M, 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. Am. J. Hum. Genet., 56: 799-810.

Long N., Gianola D., Rose G. J. M., Weigel K. A., Avendano S., 2007. Machine learning classification procedure for selection SNPs in genomic selection: application to early mortality in broilers. J. Anim. Breed. Genet. 124: 377–389.

Longmate JA, 2001. Complexity and power in case-control association studies. Am J Hum Genet., 68: 1229–1237.

Lorenz AJ, Hamblin MT, Jannink JL, 2010. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS ONE.

Lorenzana RE, Bernardo R, 2009. Accuracy of genetic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics, 120: 151-161.

Loss SP, Siddique KHM, 1994. Morphological and physiological traits associated with wheat yield increases in Mediterranean environments. Advances in Agronomy, 52: 229-276.

Lübberstedt T, 2013. Diagnostics in Plant Breeding. In: Lübberstedt T and Varshney K (Eds.), Diagnostics in Plant Breeding. Dordrecht: Springer, New York, 3-10.

Lynch M, Walsh B, 1998. Genetic and analysis of quantitative traits. In: Sunderland MA (Ed.), Sinauer Associates.

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R, 2005. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Molecular Breeding, 15: 271-290.

Maccaferri M, Sanguineti MC, Corneti S, Ortega JLA, Salem MB, Bort J, DeAmbrogio E, del Moral LFG, 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.

Mackay I, Powell W, 2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Sci, 12: 57-63.

Mackay TFC, Stone EA, Ayroles JF, 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics, 10: 565-577.

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S, 2012. SNP Markers and Their Impact on Plant Breeding. International Journal of Plant Genomics.

Mammen E, 1993. Bootstrap and wild bootstrap for high dimensional linear models. Annals of Statistics, 21: 255-285.

Manly KF, Olson JM, 1999. Overview of QTL mapping software and introduction to Map Manager QT. Mammalian Genome, 10: 327-334.

Martinez O, Curnow RN, 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theoretical and Applied Genetics, 85: 480-488.

McCarthy MI, Abecasis GR, Cardon LR, 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics, 9: 356-369.

McFadden ES, Sears ER, 1946. The origin of Triticum spelta and its free-threshing hexaploid relatives. Journal of. Heredity, 37: 81-107.

McIntyre CL, Mathews KL, Rattey A, Drenth J, Ghaderi M, Reynolds M, Chapman SC, Shorter R, 2010. Molecular detection of genomic regions associated with grain yield and yield components in an elite bread wheat cross evaluated under irrigated and rainfed conditions. Theor Appl Genet, 120: 527–541.

Metzker ML, 2010. Sequencing technologies—The next generation. Nature Reviews Genetics, 11: 31-46.

Meuwissen THE, Hayes BJ, Goddard ME, 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819–1829.

Meuwissen THE and Goddard ME, 2001. Prediction of identity by descent probabilities from marker haplotypes. Genet. Sel. Evol., 33: 605–634.

Miralles DJ, Slafer GA, 1999. Wheat development. In: Satorre EH, Slafer GA (Eds.), Wheat: Ecology and Physiology of Yield Determination. Food Production Press.

Miralles DJ, Richards RA, 2000. Response of leaf and tiller emergence and primordium initiation in wheat and barley to interchanged photoperiod. Annals Bot., 85: 655–663.

Miralles DJ, Slafer GA, 2007. Sink limitations to yield in wheat: how could it be reduced? Journal of Agricultural Science145: 1-11.

Monneveux P, Zaharieva M, Rekika D, 2000. The utilisation of Triticum and Aegilops species for the improvement of durum wheat. In: Royo C, Nachit MM, Di Fonzo N, Araus JL (Eds.), Durum Wheat Improvement in the Mediterranean Region: New Challenges. CIHEAM, Centre Udl-IRTA, ICARDA, CIMMYT, Zaragoza. Options Méditerranéennes, Series A, 40: 71-83.

Morgante M, Salamini F, 2003. From plant genomics to breeding practice. Current Opinions in Biotechnology, 14: 214-219.

Morris RW, Kaplan NL, 2002. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. Genet Epidemiol., 23: 221-233.

Mott R, Flint J, 2002. Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics, 160: 1609–1618.

Nei M, 1972. Genetic distance between populations. Am. Nat., 106: 283-292.

Nillson-Ehle H, 1909. Kreuzunguntersuchungen an Hafer und Weizen. Acta Universitatis Lundensis, 2: 1-122.

Niu T, Qin Z, Xu X, Liu JS, 2002. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. Am. J. Hum. Genet., 70: 157-159.

Nordborg M, Tavaré S, 2002. Linkage disequilibrium: What history has to tell us. Trends in Genetics, 18: 83-90.

Nyholt DR, 2004. A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. American journal of human genetics, 74: 765–769.

Ozkan H, Levy AA, Feldman M, 2001. Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops–Triticum) group. Plant Cell, 13: 1735-1747.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, et al., 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science, 294: 1719-1723.

Patterson HD, Williams ER, 1976. A new class of resolvable incomplete block designs. Biometrika 63: 83-92.

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, et al., 2004. Methods for high-density admixture mapping of disease genes. Am. J. Hum. Genet. 74: 979-1000.

Peleman JD, Rouppe van der Voort J. 2003. Breeding by design. Trends Plant Sci. 8:330–334.

Peña RJ, 2002. Wheat for bread and other foods. In: Curtis BC, Rajaram S, Macpherson HG (Eds.), Bread Wheat: Improvement and Production. FAO, Rome.

Pfeiffer WH, Sayre KD, Reynolds MP, 2000. Enhancing genetic grain yield potential and yield stability in durum wheat. Options Méditerranéennes 40: 83–93. In: C. Royo et al. (Eds.), Durum wheat improvement in the Mediterranean region: New challenges.

Piepho HP, Gauch HG, 2001. Marker pair selection for mapping quantitative trait loci. Genetics, 157: 433-444.

Pinto RS, Reynolds MP, Mathews KL, McIntyre CL, Olivares–Villegas JJ, Chapman SC, 2010. Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theoretical and Applied Genetics, 121: 1001-1021.

Pritchard JK, Stephens M, Donnelly P, 2000a. Inference of population structure using multilocus genotype data. Genetics, 155: 945-959.

Pritchard JK, 2001. Are rare variants responsible for susceptibility to complex disease? Am J Hum Genet, 69: 124-137.

Pugsley AT, 1966. The photoperiodic sensitivity of some spring wheats with special reference to the variety Thatcher. Aust J Agric Res 17: 591–599.

Pumphrey MO, Bernardo R, Anderson JA, 2007. Validating the Fhb1 QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Science, 47: 200-206.

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusić D, Waterman E, Weyen J, Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimbekova M, Turuspekov Y, Abugalieva S, Tuberosa R, Sanguineti MC, Hollington PA, Aragués R, Royo A, Dodig D, 2005. A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring x SQ1 and its use to compare QTLs for grain yield across a range of environments. Theoretical and Applied Genetics, 110: 865–880.

Rafalski A, 2002a. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol, 5: 94-100.

Rafalski JA, 2002b. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science, 162: 329–333.

Rafalski JA, 2010. Association genetics in crop improvement. Current Opinion in Plant Biology 13: 174-180.

Rakshit S, Rakshit A, Patil JV, 2012. Multiparent intercross populations in analysis of quantitative traits. Journal of Genetics, 91: 111-117.

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, et al., 2006. Single-nucleotide polymorphisms (SNPs) frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome, 49: 1131-1139.

Rebai A, Goffinet B, Mangin B, 1995. Comparing powers of different methods for QTL detection. Biometrics, 51: 87-99.

Rebai A, Goffinet B, 2000. More about quantitative trait locus mapping with diallel designs. Genome Research, 75: 243-247.

Rebetzke GJ, Condon AG, Rattey AR, Farquhar GD, Richards RA, 2012. Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in bread wheat (Triticum aestivum L.). Functional Plant Biology, 40: 14-33.

Remington DL, Whetten RW, Liu B, O'Malley DM, 1999. Construction of an AFLP genetic map with nearly complete genome coverage in Pinus taeda. Theor Appl Genet, 98: 1279-1292.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Science, 98: 11479–11484.

Reynolds MP, Trethowan RT, van Ginkel M, Rajaram S, 2001. Application of physiology to wheat breeding. In: Reynolds MP, Ortiz-Monasterio I, McNab A (Eds.), Application of physiology in wheat breeding. Mexico, DF, CIMMYT.

Reynolds MP, Pellegrineschi A, Skovmand B, 2005. Sink-limitation to yield and biomass: a summary of some investigations in spring wheat. Annals of Applied Biology, 146: 39-49.

Reynolds MP, Calderini DF, Vargas M, 2007. Association of source/sink traits with yield, biomass and radiation use efficiency among random sister lines from three wheat crosses in a high -yield environment. Journal of Agricultural Science, Cambridge 145: 3-16.

Reynolds M, Tuberosa R, 2008. Translational research impacting on crop productivity in drought-prone environments. Curr. Opin. Plant Biol., 11: 171–179.

Reynolds M et al, 2009. Raising yield potential in wheat. J. Exp. Bot., 60: 1899–918.

Richards RA, 1992a. The Effect of Dwarfing Genes in Spring Wheat in Dry Environments. I. Agronomic Characteristics. Spring.

Richards RA, 1992b. The Effect of Dwarfing Genes in Spring Wheat in Dry Environments. II.* Growth, Water Use and Water-use Efficiency. Water.

Rincent R, Laloe D, Nicolas S, Altmann T, Brunel D, et al., 2012. Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (Zea mays L.). Genetics, 192: 715-728.

Risch N, Merikangas K, 1996. The future of genetic studies of complex human diseases. Science, 273: 1516-1517.

Robertson MJ, Brooking IR, Ritchie JT, 1996. Temperature response of vernalization in wheat: modelling the effect on the final number of mainstem leaves. Annals of Botany, 78: 371-381.

Rockman MV, Kruglyak L, 2008. Breeding designs for recombinant inbred advanced intercross lines. Genetics, 179: 1069-1078.

Rodriguez VM, Butron A, Malvar RA, Ordas A, Revilla P, 2008. Quantitative trait loci for cold tolerance in the maize IBM population. Int. J. Plant Sci. 169: 551-556.

Sakamoto T and Matsuoka M, 2008. Identifying and exploiting grain yield genes in rice. Curr. Opin. Plant Biol. 11: 209–214.

Sadras VO and Lawson C, 2011. Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. Crop Pasture Sci. 62: 533–549.

Sarkar P, Stebbins GL, 1956. Morphological evidence concerning the origin of the B genome in wheat. American Journal of Botany, 43: 297-304.

Sax K, 1923. The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics, 8: 552-560.

Schefers, J.M. and K.A. Weigel. 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Animal Frontiers 2: 4-9.

Semagn K, Bjørnstad Å, Ndjiondjop MN, 2006. Progress and prospects of marker assisted backcrossing as a tool in crop breeding programs. African Journal of Biotechnology, 5: 2588-2603.

Semagn K, Bjørnstad Å, Xu Y, 2010. The genetic dissection of quantitative traits in crops. Electron J Biotechnol 13: 5.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA, 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell, 13: 1749-1759.

Siddique KH, Kirby EJM, Perry MW, 1989. Ear stem ratio in old and modern wheat varieties: relationship with improvement in number of grains per ear and yield. Field Crops Research, 21: 59-78.

Skovmand B, Peynold MP, Delacy IH, 2000. Mining wheat germplasm collection for yield enhancing traits. Proceeding of the 6th International Wheat Congress, June 5-9, 2000. Budapest, Hungary.

Slafer GA and Andrade FH, 1989. Genetic improvement in bread wheat (*Triticum aestivum*, L.) yield in Argentina. Field Crops Res. 21: 289–296.

Slafer GA, Andrade FH, 1993. Physiological attributes related to the generation of grain yield in bread wheat cultivars released at different eras. Field Crops Research, 31: 351-367.

Slafer GA and Rawson HM, 1994. Sensitivity of wheat phasic development to major environmental factors: a re-examination of some assumptions made by physiologist and modellers. Australian Journal of Plant Physiology, 21: 393-426.

Slafer GA, Satorre EH, Andrade FH, 1994. Increases in grain yield in bread wheat from breeding and associated physiological changes. In: Slafer GA (Ed.), Genetic improvement of field crops. New York: Marcel Dekker Inc., 1-68.

Slafer GA, Calderini DF, Miralles DJ, 1996. Yield components and compensation in wheat: opportunities for further increasing yield potential. In: Reynolds MP, Rajaram S, McNab A (Eds.), Increasing Yield Potential in Wheat: Breaking the Barriers. Reynolds, M.P., S. Rajaram, and A.McNab, eds., Mexico, D.E: CIMMYT.

Slafer GA, 2003. Genetic basis of yield as viewed from a crop physiologist's perspective. Annals of Applied Biology, 142: 117-128.

Slafer GA, Calderini DF, 2005. Importance of breeding for further improving durum wheat. In: Royo et al. (Eds.), Durum wheat breeding. Food Products Press, NewYork, Vol. I. 87-89.

Slager SL, Huang J, Vieland VJ, 2000. Effect of allelic heterogeneity on the power of the transmission disequilibrium test. Genet. Epidemiol., 18: 143-156.

Slageren M, 1994. Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). ICARDA / Wageningen Agricultural University Papers, 94(7): i-xiv, 1-512.

Soller M, Brody T, Genizi A, 1976. On the power of experimental designs for detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theoretical and Applied Genetics, 47: 35-39.

Soller M, Beckmann JS, 1990. Marker-based mapping of quantitative trait loci using replicated progenies. Theoretical and Applied Genetics, 80: 205-208.

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A, 2003. Mining single nucleotide polymorphisms from hexaploid wheat ESTs. Genome, 49: 431-437.

Somers DJ, Isaac P and Edwards K, 2004. A high density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet, 109:1105-1114.

Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C, 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome, 50: 557-567.

Sourdille P, Tixier MH, Charmet G and Gay G, 2000. Location of genes involved in ear compactness in wheat (*Triticum aestivum*) by means of molecular markers. Mol. Breed. 6: 247-255.

Sourdille P, Sukhwinder T, Cadalen G, Brown-Guedira G, Gay et al., 2004. Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). Funct. Integr. Genomics 4: 12–25.

Stephens M, Smith NJ, Donnelly P, 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet., 68: 978-989.

Talbert LE, Blake NK, Storlie EW, Lavin M, 1995. Variability in wheat based on low-copy DNA sequence comparisons. Genome National Research Council of Canada, 38: 951-957.

Talbot CJ, Nicod A, Cherny SS, Fulker DW, Collins AC, Flint J, 1999. High-resolution mapping of quantitative trait loci in outbred mice. Nat Genet, 21: 305-308.

Tanksley SD, 1993. Mapping polygenes. Annual Review of Genetics, 27: 205-233.

Teuscher F, Broman KW, 2007. Haplotype probabilities for multiple-strain recombinant inbred lines. Genetics, 175: 1267-1274.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES, 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics, 28: 286-289.

Threadgill DW, Hunter KW, Williams RW, 2002. Genetic dissection of complex and quantitative traits: From fantasy to reality via a community effort. Mamm Genome, 13: 175–178.

Tranquilli GE, Dubcovsky J, 2000. Epistatic interactions between vernalization genes Vrn-Am1 and Vrn-Am2 in diploid wheat. Journal of Heredity, 91: 304-306.

Trevaskis B, Hemming MN, Peacock WJ, Dennis ES, 2007. The molecular basis of vernalization-induced flowering in cereals. Trends in Plant Science, 12: 352-357.

Trick M, Long Y, Meng J, Bancroft I, 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnology, 7: 334-346.

Turner A, Beales J, Faure S, Dunfordand RP, Laurie DA, 2005. The pseudo-response regulator *Ppd*-H1 provides adaptation to photoperiod in barley. Science, 310: 1031-1034.

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J, 2006. A *NAC* gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science, 314: 1298-301.

Uleberg E, Meuwissen THE, 2007. Fine mapping of multiple QTL using combined linkage and linkage disequilibrium mapping – A comparison of single QTL and multi QTL methods. Genet Sel Evol, 39: 285-299.

Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J, 2006b. Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet., 38: 879-887.

Valente F, Gauthier F, Bardol N, Blanc G, Joets J, Charcosset A, Moreau L, 2013. OptiMAS: a decision support tool for marker-assisted assembly of diverse alleles. J Hered. 104(4): 586-90.

van Eeuwijk FA, Boer M, Totir LR, Bink M, Wright D et al., 2010. Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. Theor Appl Genet, 120: 429-440.

van Slageren MW, 1994. Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. and Spach) Eig (Poaceae). Agricultural University, Wageningen & International Center for Agricultural Research in Dry Areas, Aleppo, Syria.

Varshney RK, Hoisington DA, Nayak SN, Graner A, 2009. Molecular plant breeding: methodology and achievements. In: Somers D, Langridge P, Gustafson PJ (Eds.), Methods in molecular biology: Plant genomics. Totowa, NJ: The Humana Press, 283-304.

Vorrips RE, 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity, 93: 77–78.

Wang N, Akey JM, Zhang K, Chakrabortyand R, Jin L, 2002. Distribution of recombination crossovers and the originof haplotype blocks: the interplay of population history, recom-bination, and mutation. Am. J. Hum. Genet., 71: 1227-1234.

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, 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 AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E, 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J., In press.

Whittaker JC, Curnow RN, Haley CS, Thompson R, 1995. Using marker-maps in marker-assisted selection. Genetical Research, 66: 225-232.

Whittaker JC, Thompson R, Denham MC, 2000. Marker-assisted selection using ridge regression. Genetics Research, 75: 249-252.

Wilhelm EP, Turner AS, Laurie DA, 2009. Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (*Triticum durum* Desf.). Theoretical and Applied Genetics, 118: 285-294.

Winkler CR, Jensen NM, Cooper M, Podlich DW, Smith OS, 2003. On the determination of recombination rates in intermated recombinant inbred populations. Genetics, 164: 741-745.

Worland AJ, Law CN, 1986. Genetic analysis of chromosome 2D of wheat. I 1. The location of genes affecting height, day-length insensitivity, hybrid dwarfism and yellow-rust resistance. Zeitschrift für Pflanzen, 96, 331–345.

Würschum T, 2012. Mapping QTL for agronomic traits in breeding populations. Theoretical and Applied Genetics, 125: 201-210.

Xu S, 1995. A comment on the simple regression method for interval mapping. Genetics, 141: 1657-1659.

Xu Y, Crouch JH, 2008. Marker-assisted selection in plant breeding: from publications to practice. Crop Science, 48: 391-407.

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, Shibaya T, Yano M, 2010. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11: 267.

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, 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.

Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J, 2004. Allelic variation at the *VRN-1* promoter region in polyploid wheat. TAG. Theoretical and Applied Genetics, 109: 1677-86.

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.

Youssefian S, Kirby EJM, Gale MD, 1992. Pleiotropic effects of the GA-insensitive *Rht* dwarfing genes in wheat 2. Effects on leaf, stem, ear and floret growth. Field Crops Research, 28: 191-210.

Yu J, Buckler ES, 2006. Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17: 155-160.

Zadoks JC, Chang TT, Konzak CF, 1974. A decimal code for the growth stages of cereals. Weed Research, 14: 415-421.

Zanetti S, Winzeler M, Feulillet C, Keller B, Messmer M, 2001. Genetic analysis of bread-making quality in wheat and spelt. Plant Breed, 120: 13-19.

Zeng Z, 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proceedings of the National Academy of Sciences, 90: 10972-10976.

Zeng Z, 1994. Precision mapping of quantitative trait loci. Genetics, 136: 1457-1468.

Zhang LY, Liu DC, Guo XL, Yang WL, Sun JZ, Wang DW, Zhang A, 2010. Genomic Distribution of Quantitative Trait Loci for Yield and Yield-related Traits in Common Wheat. Journal of integrative plant biology, 52(11): 996-1007.

Zhu C, Gore M, Buckler ES, Yu J, 2008. Status and prospects of association mapping in plants. Plant Genome 1**:** 5–20.