Justus Liebig University, Giessen

Institute of Plant Breeding and Agronomy I

Department of Plant Breeding

Analysis and use of genomic diversity in hexaploid wheat (*Triticum aestivum* L.)

Inaugural Dissertation for a Doctorate Degree in Agricultural Sciences

in the Faculty of Agricultural Sciences, Nutritional Sciences and

Environmental Management

Examiners

1. Prof. Dr. Rod J. Snowdon
2. Prof. Dr. Matthias Frisch

Submitted by

Kai Peter Voss-Fels

Giessen 2016

„Alles kommt in der Wissenschaft auf das an, was man ein Aperçu nennt, auf ein Gewahrwerden dessen, was eigentlich den Erscheinungen zum Grunde liegt. Und ein solches Gewahrwerden ist bis ins Unendliche fruchtbar.“

(***Johann Wolfgang von Goethe***, Naturwissenschaftliche Schriften, 1784-1810.
Zur Geschichte der Farbenlehre. Galileo Galilei)

Table of Contents

1 General introduction

1.1 The role of wheat in a dynamic world

Bread wheat (*Triticum aestivum* L.), the staple food for one third of the global population, is one of the most important crop plants worldwide. In 2014, almost 730 M t seeds were produced on 220 M ha worldwide, making it one of the three most important cereal crops in the world alongside rice (740 M t in 2014) and maize (912 M t in 2014) (FAOSTAT 2016) (http://faostat.fao.org). The seeds of wheat have a particularly high starch content which enables various uses in human nutrition. Wheat grain is also used for animal feeding and, more recently, for bioethanol production (Balat et al., 2008; Shewry, 2009). In 2014, the worldwide average yield was 3.6 t/ha, however there is a huge global variation in production volume per area. This is mainly caused by regional environmental conditions, differences in farming practices and different eco-geographical forms of wheat grown in different climatic zones. For example, Australian farmers grow mostly spring-sown wheat with a short vegetation period, on water-limited sites; their 2014 harvest was consequently 1.6 t/ha. In contrast, average yields over six times higher (~9.9 tons/ha) were achieved in 2014 by top European wheat producers in Germany or the Netherlands, where mainly winter wheat is grown, with a prolonged vegetative period, under optimal conditions (FAOSTAT 2016).

While the world's population is predicted to expand to more than nine billion people by the middle of this century, the available agricultural area on which crops can be produced will increase to a much lower extent, if at all. This necessitates substantial yield increases of all major commodity crops, with predictions suggesting that up to 40 % more production will be necessary (Spiertz and Ewert, 2009) in order to meet the demand for 70 % more food, as predicted by the declaration of the World Summit on Food Security in 2009 (Grainger, 2010). The need to increase the efficiency of crop production systems for global major players in food supply, such

as wheat, represents a serious challenge for the agricultural and environmental research community.

At the same time, wheat yield increases have stagnated during the last two decades in many parts of the world, including Europe, the United States of America, Asia and Australia (Brisson et al., 2010; Ray et al., 2012). In Europe, the largest wheat producer worldwide, this stagnation is attributed to a drastic reduction of genetic diversity within elite wheat breeding gene pools, during the course of strong selective breeding and intensive germplasm exchange between different commercial breeding programs (Reif et al., 2005). On the other hand, global wheat production is threatened by an increasing frequency of extreme environmental conditions in the course of climate change. It has recently been estimated that global wheat yields will be reduced by 6 % for each 1 °C of further temperature increase, accompanied by increasingly variable yield stability (Asseng et al., 2015). A recent study, comparing the influence of different extreme weather disasters on global crop production from 1964-2007, concluded that droughts and extreme heat have had the most serious negative impact on agricultural production, causing average national crop failures of 9-10 % (Lesk et al., 2016). Interestingly, the same study found that recent extreme droughts had a significantly larger effect on cereal production than earlier ones, whereas no significant effects were found due to floods or extreme cold temperature events. Furthermore, in regional and crop-specific analyses it was concluded that recent droughts have had an 8-11 % more severe effect in developed countries compared to developing countries.

Although there are inevitable projection uncertainties, droughts are very likely to become more frequent in future (Seneviratne et al., 2012), while extreme heat events are also projected to be increasingly common and severe (Battisti and Naylor, 2009). Given these extremely dynamic demographic and environmental global changes, future wheat production faces novel and

unprecedented threats that address all related fields of public and private research. An important solution in this context is the reinstitution and/or creation of wheat cultivars with broad diversity to meet specific demands of local adaptation to new climatic challenges. The genetic improvement of modern varieties to enhance climate resilience is therefore a key aspect to secure future food supply. The urgency and importance has recently also been recognized by G20 policymakers, who approved the initiation of the International Research Initiative for Wheat Improvement (IRIWI), a huge global collaborative research project aimed at optimization and strengthening of public and private wheat research endeavours (G20, 2011).

1.2 The biology of wheat and its genomic complexity

Wheat is the most widely grown cereal crop genus worldwide and cultivated in various eco-geographical regions. Like all other cereals it belongs to the *Poaceae* family, which includes monocotyledonous flowering plants colloquially known as 'grasses'. Around ten thousand years ago, along with the first agricultural settlements of the Fertile Crescent, farmers began to make use of wild diploid wheat species (e.g. *Triticum* and *Aegilops* species) (Marcussen et al., 2014). As agriculture evolved these were gradually replaced by domesticated diploids and newly formed, allopolyploid wheat forms (Salamini et al., 2002). Today, the allohexaploid bread wheat (*Triticum aestivum* L., $2n = 6x = 42$ chromosomes, genomic composition AABBDD) clearly dominates global wheat production. Bread wheat is a spiked annual grass that grows between 0.8 and 1.5 m high. It produces a very fine root system, with three radicles that can penetrate the soil to a depth of around one meter. A crop with moderate to high input demands, wheat requires heavy clay soils with an optimal nutrient supply and a high water storage capacity. Hexaploid wheat, an autogamous self-pollinator with hermaphrodite flowers and an outcrossing rate of less

than 5%, bears spikes consisting of several spikelets, arranged linearly along the rachis. Each spikelet contains 2-5 florets that produce 2-4 kernels each. Due to the global widespread of wheat cultivation and the local adaptation of regionally grown cultivars, there is a broad diversity of growth requirements among the various varieties, for example for day length or vernalization, the plant's requirement for prolonged cold exposure as a prerequisite for flowering (Deng et al., 2015).

Besides its outstanding global economic importance, bread wheat distinguishes itself from other crops by its enormous (~17,000 Mbp), complex, allohexaploid genome. The genome consists of three independent subgenomes (A, B and D -subgenomes) derived from three ancestral hybridization events (Fig. 1). Initially, the A and B-subgenome donors, diploid species of the genera *Triticum* and *Aegilops*, diverged into two independent lineages from a common ancestor around 6.5 Mio years ago. These lineages subsequently formed the D-subgenome donor in the first major hybridization event. In the second event that gave rise to allotetraploid emmer wheat (AABB), the A-subgenome donor *T. urartu* hybridized with a close relative of the B-subgenome donor *Ae. speltoides*. Finally, emmer wheat (AABB) and *Ae. tauschii* (DD) formed today's bread wheat (AABBDD) in the third major hybridization event (Marcussen et al., 2014). Today, most ancient wheat species, almost completely been replaced in agricultural usage by modern bread wheat, are cryopreserved in gene banks in order to maintain this highly valuable resource of genetic diversity. In total, there are over 560,000 wheat accessions stored in around 40 gene banks worldwide (Bhullar et al., 2009). To achieve further genetic improvement of modern cultivars and to meet the demands of a rapidly changing world, the detailed genomic exploitation of these ancient species is assumed to be a key strategy (Longin and Würschum, 2016).

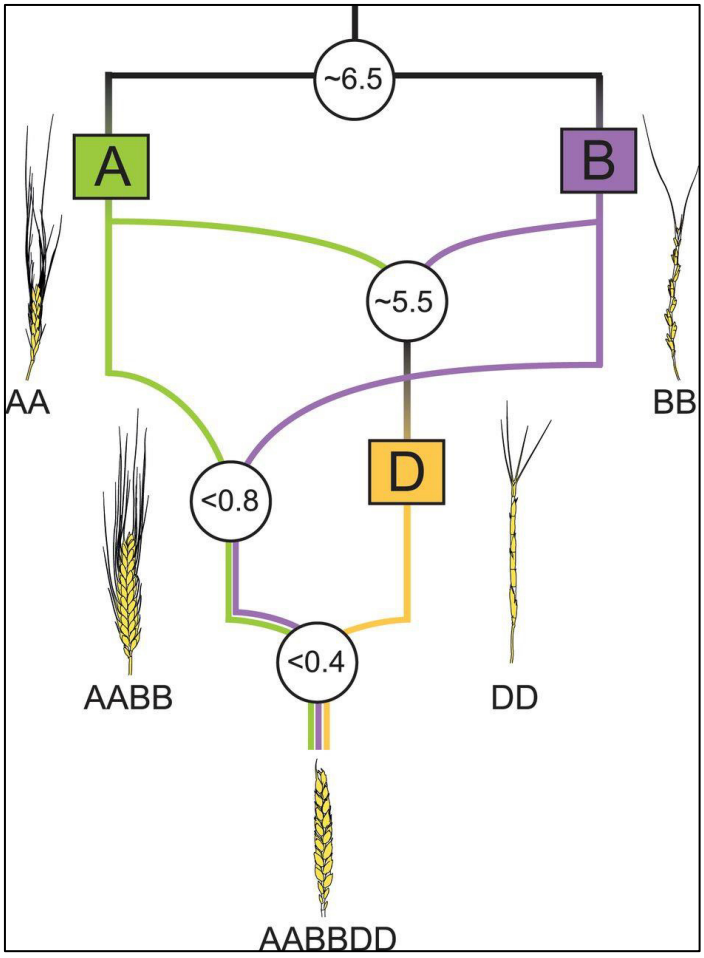


Figure 1. Evolution of allohexaploid wheat (*T.* aestivum). Approximate dates for divergence and the three hybridization events are given in white circles in units of million years ago (Mya). Differentiation of the wheat lineage (*Triticum* and *Aegilops*) from a common ancestor into the A and B genome lineages began ~6.5 Mya. The first hybridization occurred ~5.5 Mya between the A and B genome lineages, and also led to the origin of the D genome lineage by homoploid hybrid speciation. The second hybridization, between a close relative (BB) of *Ae. speltoides* and *T. urartu* (AA), gave rise to the allotetraploid emmer wheat (*T. turgidum*; AABB) by allopolyploidization. Bread wheat originated with a further allopolyploidization event following a third hybridization, between emmer wheat and *Ae. tauschii* (DD). The three diploid lineages are indicated with color and labels. Images of inflorescences (spikes) illustrate extant species closely related to those involved in the polyploidizations. (Source: Marcussen et al., 2014).

With the rapid development of modern biotechnology and sequencing tools, genomics-based crop research that allows large-scale, high-resolution population characterization has been

revolutionized during the last two decades. The major prerequisite for genomics-based exploitation of valuable diversity is a deep genome investigation and the generation of high-quality reference genome sequences. It is now over 15 years since the first plant genome sequence of *Arabidopsis thaliana* was fully decoded (genome size: 125 Mbp) (The Arabidopsis Genome Initiative, 2000), shortly followed afterwards by rice (430 Mbp) (Goff et al., 2002). Today, the fast development of next-generation sequencing (NGS) platforms lay the foundation for reference sequences of complex and large genomes from almost all important crop species, including the major cereals maize (2,500 Mbp) (Schnable et al., 2009), barley (5,100 Mbp) (Mayer et al., 2012) and sorghum (730 Mbp) (Paterson et al., 2009). Deep wheat genome analyses are hampered by the extremely large genome, the very high fraction of non-coding, repetitive DNA (>80%) and the close relatedness of the A, B and D subgenomes (Gupta et al., 2008). Recently, a combination of next-generation sequencing technologies with flow cytometry and syntenic mapping enabled the establishment of a first chromosome-based draft genome sequence for hexaploid wheat (The International Wheat Genome Sequencing Consortium (IWGSC), 2014). In June 2016, the preliminary assembly of a first full reference genome sequence was made available by the International Wheat Genome Sequencing Consortium (IWGSC) to scientists worldwide, and a "gold standard" completed reference sequence is expected to be released in 2017 (http://www.wheatgenome.org). At the same time, ultrafast reduced-representation DNA sequencing technologies have paved the way for a rapid molecular marker discovery by resequencing of diverse germplasm collections, even in complex genomes like that of wheat (Edwards et al., 2013). Such datasets have also enabled the introduction of commercial high-throughput genotyping platforms, harboring up to hundreds of thousands of single-nucleotide-polymorphism (SNP) markers. This provided an effective and straight-forward means for various genetic research purposes (Ganal et al., 2012). These technologies have also

achieved great importance in modern crop breeding and are undergoing continual further efficiency and cost improvements (Poland and Rife, 2012). Today, there are high-capacity SNP arrays available for all major cereal crops including maize (50–600K SNP arrays) (Ganal et al., 2011; Unterseer et al., 2014), rice (51.5K SNP array) (Chen et al., 2014) sorghum (90K SNP array) (Wieckhorst et al., 2015) and wheat (9, 35, 90 and 800K SNP arrays) (Cavanagh et al., 2013; Wang et al., 2014; Winfield et al., 2016).

1.3 Breeding in the genomics era: The impact of modern tools on wheat improvement

Conventional bread wheat breeding has a long history of success in the improvement of yield, quality and resistance related traits (Baenziger et al., 2014). Since the earliest documentation of wheat yields by the Food and Agriculture Organization of the United Nations (FAO) in the 1960s, global average yields could gradually be increased, from less than 1.1 t/ha up to 3.6 t/ha in 2014 (FAOSTAT 2016). The most spectacular yield increases in the history of wheat production, achieved during the "Green Revolution" in the 1960s, can largely be attributed to the introgression of different dwarfing genes ('reduced height', *Rht* genes) into modern high-yielding cultivars. These genes affect gibberellin acid (GA), a crucial plant hormone which causes a significant shortening of the stem. The considerably more stable stems of shorter wheat lines are capable of carrying heavier spikes, with increased numbers of kernels, without simultaneously becoming more prone to lodging (Hedden, 2003). Specific end-use characteristics, including the glutenin subunit composition that mainly affects baking quality, have continuously been improved (Payne et al., 1987) and methods that allowed an indirect selection of quality characteristics based on seed protein profiles were an early development that facilitated breeding progress (Koebner and Summers, 2003). Further breeding success could be achieved by the

establishment of vastly improved disease resistances, particularly against powdery mildew and different species of rust pathogens, by combining different resistance genes in modern cultivars (Ellis et al., 2014).

Classically, wheat varieties are produced in line breeding approaches that are common for self-fertilizing crops, mostly via bulk or pedigree selection. In the simplest form of the pedigree breeding method, two parental genotypes are crossed and produce a uniform F1 population (Fig. 2). Starting from the first segregating F2 generation, the breeder directly selects single plants based on their characteristics for traits with monogenic and oligogenic inheritance, such as resistances or quality traits. From the F6 generation onwards, when genotypes are assumed to have a high level of homozygosity (~97%) yield trialling is initiated. The generation time for winter wheat is normally a full calendar year, thus the registration of a new variety that has been bred by conventional pedigree breeding generally takes between 10-15 years. The invention of molecular DNA markers and the consequential concept of indirect marker-assisted selection (MAS) were expected to improve this tedious breeding process by drastically reducing time and costs. After the development of first marker technologies, initially restriction fragment length polymorphism (RFLP) and early PCR-based methods like amplified fragment length polymorphism (AFLP) markers, the relatively simple and cost-effective use of simple sequence repeat (SSR, or 'microsatellite') markers became the method of choice in cereal genomics (Collard and Mackill, 2008). Microsatellites are repetitive tracts of short DNA sequences, between 2-5 base-pairs long, that are highly polymorphic between individuals and easily amplified by PCR with primers based on DNA sequences flanking the SSR.

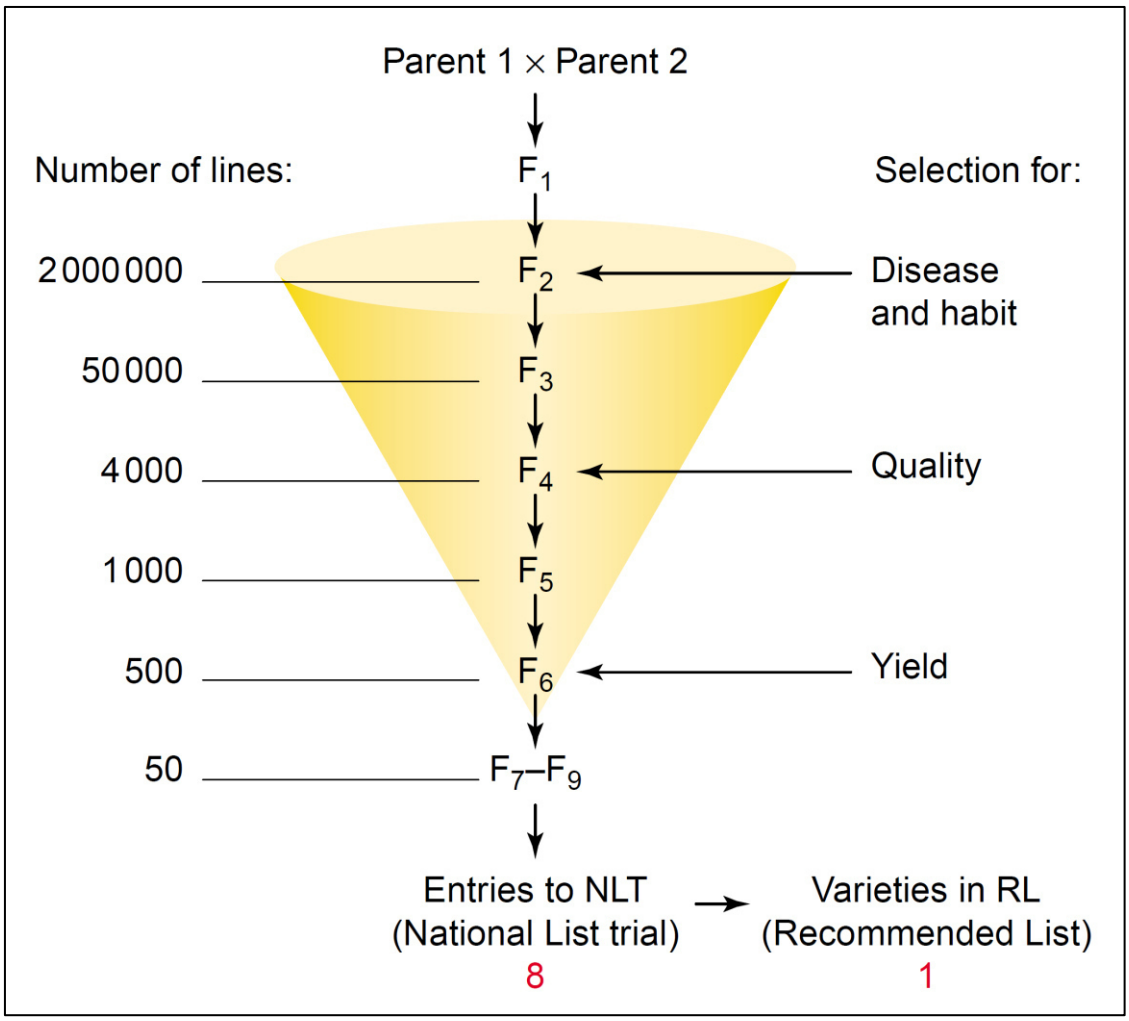


Figure 2. The selection funnel in a pedigree breeding program for wheat. Representative numbers of selections taken forward in each generation are shown along the left edge of the triangle. The generation time for winter wheat is typically a full calendar year. (Source: Koebner and Summers, 2003).

Despite the advantages of SSR markers regarding reproducibility, reliability and their ability to capture co-dominant inheritance, their analysis requires electrophoresis-based methods, so that multiplexing, for simultaneous testing of high numbers of markers, is strongly limited. Nevertheless, there are numerous genetic studies in wheat that describe the application of SSR markers, for example in the characterization of breeding material (Chen et al., 2012; Fu et al.,

2006; Manifesto et al., 2001; Roussel et al., 2005) or in the identification of quantitative trait loci (QTL) for various traits including fusarium head blight (FHB), scab resistance (Anderson et al., 2001; Zhou et al., 2003), protein content (Prasad et al., 2003) or yield-related traits (Li et al., 2007). Due to their low cost and simplicity, many small-scale commercial wheat breeding operations still implement SSR markers, particularly for marker-assisted backcrossing to select for important recessive, monogenic resistances.

Today, modern high-throughput genotyping tools, which potentially enable the rapid characterization of millions of SNPs, insertion/deletion polymorphisms (InDels) or copy-number variants (CNV), have almost entirely replaced classical marker systems like SSRs for many important crops. A variety of high-throughput genotyping platforms are used in modern wheat breeding research, including SNP arrays from the market-leaders in the array sector, Illumina Inc. (San Diego, CA) and Affymetrix Inc. (Santa Clara, CA) with 15K, 35K, 90K or 850K SNPs (Bassi et al., 2016, 2016; Wang et al., 2014; Winfield et al., 2016). Other options available include diversity array technology (DArT) marker arrays, which can either detect SNPs or InDels (Akbari et al., 2006), and genotyping-by-sequencing (GBS) approaches based on next-generation sequencing of restriction fragment ends (Poland, 2015). Table 1 summarizes different genotyping platforms that are currently in use for various uses in wheat. The possible applications of large genomic data produced by novel high-throughput systems in genetic studies are various. They include assessment of population structure and genotypic diversity on a genome-wide, subgenome-wide or chromosome-wide scale, identification of selective sweeps and signatures of directional selection, targeting genetic variants that are associated with agronomic traits, genomic selection (GS) and the prediction of hybrid performance. High-throughput genotyping has further

been used in approaches that aim to unravel the genetic basis of heterosis (Voss-Fels and Snowdon, 2016).

Table 1. Comparison of genotyping platforms in wheat.

Platform	Marker No.[a]	Cost per individual (US $)[b]	Advantage	Disadvantages	Provider
KASPAr	90	1	No missing data Co-dominant	Bi-allelic	LGC Genomics
Illumina 15K	10,000	35	No missing data Co-dominant	Bi-allelic	Various
Genotyping-by-sequencing (GBS)	10,000	12	Multi-allelic Co-dominant	Considerable missing data Massive data	Universities
Diversity array technology (DArT)	10,000	25	Multi-allelic Co-dominant	Missing data	Triticarte
Illumina 90K	15,000	50	No missing data Co-dominant	Bi-allelic	Various
Axiom 35K	20,000	50	No missing data Co-dominant	Bi-allelic	Affymetrix
Axiom 850K	300,000	250	No missing data Co-dominant	Bi-allelic Massive data	Affymetrix

[a]Expected number of polymorphic markers based on literature.
[b]Derived from quotes received in the past 12 months; each provider might change the price based on the population size or established collaborations. These estimates do not include the cost of DNA extraction nor of shipping plates to providers. (Source: Bassi et al., 2016)

Turuspekov et al. (2015) used the 90K wheat Illumina SNP array to perform a phylogenetic analysis of wheat lines from Kazakhstan. Based on the pairwise genetic relationships between the genotypes that were calculated with the genome-wide marker data, they were able to show that

groups of wheat lines from the U.S.A. show high genetic similarities with different Kazakh wheat genotypes. Another recent study investigated the structural diversity among a total of 42,138 diverse hexaploid wheat lines from the national small grain collection of the United States Department of Agriculture Agricultural Research Service (USDA-ARS). That study analyzed a core collection consisting of almost 3,300 accessions with over 5,000 molecular markers (Bonman et al., 2015), mostly from the 9k SNP array (Cavanagh et al., 2013). By performing an analysis of molecular variance, principal coordinate analysis, cluster analysis and by ranking the contribution of single lines to the overall diversity, it was found that the revealed genetic subgroups corresponded well to their geographic origin and that most of the population stratification was accounted for by differences between Iranian wheat lines and the rest of the accessions. This work furthermore provided a basis for the establishment of different core subsets from the huge USDA-ARS germplasm collection. In a study by Cavanagh et al. (2013) they analyzed 2,994 worldwide hexaploid wheat accessions including landraces and modern varieties with their newly developed 9k SNP array. By using the genome-wide SNP data, clear genetic subgroups corresponding to ecogeographic origin could be identified, revealing general and subgroup-specific hotspots of directional selection, so-called ‘selective sweeps’ that most likely evolved in the course of local adaptation and strong selection for specific genetic variants during crop evolution and improvement. Affected genes were associated with plant developmental characteristics, such as inflorescence development or phenology. Strong genetic fixation was also found around a gene conferring fungal resistance. This work sheds light on ecogeographically specific genetic effects of crop improvement associated with adaptation, helping to identify conserved hotspots in the wheat genome under strong directional selection. These are ideal genomic targets for a marker-based targeted introgression of novel diversity to broaden overall phenotypic variation as the basis for selection in breeding programs.

In recent years, numerous genome-wide association studies (GWAS) using high-throughput marker technologies were conducted to unravel the complex genetic architectures of various quantitatively inherited traits and to identify trait-associated molecular variants than can be used for genomics-based selection in wheat improvement. For example, Joukhadar et al. (2013) performed a genome-wide association mapping approach with 134 wheat lines and 2,518 DArT markers to identify genomic regions associated with resistances to the five major wheat pests Hessian fly, Russian wheat aphid, Sunn pest, wheat stem saw fly and cereal leaf beetle. A total of 26 significant marker-trait-associations were identified, of which 20 were novel QTL that had not previously been described. Using an *in silico* approach they could narrow down the list of genes to a few candidates that were likely to be directly involved in disease response. Zanke et al. (2015) used a combination of markers from the 90K Illumina wheat SNP array, along with 732 genome-wide SSR markers, to localize QTL for thousand grain weight in a panel of 358 European winter wheat lines. By identifying over 300 significant marker-trait associations, for which only three SNPs explained more than 6% of the phenotypic variance, the quantitative inheritance of this trait could be elucidated. Arruda et al., (2016a) used a reduced-representation GBS approach to genotype a population of 273 U.S. wheat lines, yielding a total of 19,992 polymorphic SNPs that densely covered all 21 wheat chromosomes. They subsequently tested all lines for FHB resistance and found ten highly significant, novel marker-trait associations on different chromosomes. Tagging of the most effective resistance gene for this severe wheat disease, *Fhb1*, with multiple SNPs, provided an improved opportunity for a marker-based enrichment of favorable resistance alleles into new breeding lines.

The major limitation for incorporating findings from QTL mapping or GWAS into marker-assisted crop improvement is that both detect only trait-associated molecular variants that exceed

a certain statistical threshold (Korte and Farlow, 2013). However, many agronomically important traits such as yield are not mono- or oligogenically inherited, but rather controlled by a multitude of genes with only small effects on the phenotype. Hence, in all but very few cases single-marker selection could not meet the high expectations that were held for improving breeding progress for complex traits (Baenziger et al., 2014). To overcome this problem, the concept of genomic selection (GS) with densely-spaced genome-wide markers is presently being adopted for many crop breeding programs. This method, initially conceived for animal breeding in 2001 (Meuwissen et al., 2001), is based on the concept of predicting the breeding value of an individual that has not been phenotyped purely on the basis of its genome-wide marker profile. To achieve this, a statistical prediction model needs to be trained, with phenotype and genome-wide genotype data of a so-called 'training population', in order to estimate the individual genetic contribution of every single marker to the trait of interest. Ultimately, the performance of an untested individual can then be scored by summing up the allele effects of all markers as a genomic estimated breeding value (GEBV). Several recent studies have validated the applicability of GS in wheat improvement. For example, Liu et al. (2016) showed that GS for baking quality-related traits greatly outperformed MAS and could be a powerful tool to accelerate progress in selecting for baking quality in modern European bread wheat breeding programs. Battenfield et al. (2016) confirmed these findings by implementing GS in a large wheat breeding program of the International Maize and Wheat Improvement Center (CIMMYT) in Mexico, including almost 8,000 wheat lines and over 20,000 SNPs per line, generated by GBS. That study found GS to be a powerful tool to facilitate selection for end-use wheat quality in early generations. GS was further demonstrated to be effective for improving stem rust and FHB resistance (Arruda et al., 2016b; Rutkoski et al., 2014) or yield (Poland et al., 2012). However, additional work still needs to be done in order to further validate the applicability of GS for

important traits in commercial breeding programs, and to elaborate breeding designs in order to maximize profits obtained by GS. At present, the true gain from GS in commercial wheat breeding remains elusive and studies to date indicate that the optimal GS breeding design depends on the breeder's individual needs (Bassi et al., 2016).

1.4 Prospects and future challenges for wheat production and breeding

Besides the achievements in the genetic improvement of traits related to yield, quality and resistance in modern varieties, the global security of future wheat production for food supply remains challenging. Yield increases in major wheat growing regions including Europe, Asia, Australia and the US are currently stagnating (Ray et al., 2012). One future opportunity to boost wheat productivity and drastically enhance yields in the near future is seen in hybrid breeding approaches, which already laid the foundation for a more than five-fold yield increase in maize yield during the past century (Marulanda et al., 2016). Currently, different factors hamper the successful implementation of hybrid breeding approaches in wheat. First, large-scale hybrid seed production is complicated, due to the strong self-pollinating nature of wheat (<5% outcrossing) and its hermaphrodite inflorescence. This can be overcome using chemical hybridization agents (CHA) or with genetic mechanisms that induce male sterility (e.g. cytoplasmic male sterility, CMS), in order to ensure a controlled fertilization between crossing lines. CHAs are mainly hindered by a very narrow, weather-dependent application window, while present wheat CMS systems lack stability. Furthermore, the only currently available CMS source, from *T. timopheevii*, also conveys environment-dependent, negative side-effects. Identifying and incorporating more effective sources of genetic male sterility, combined with a modification of the flower structure to enhance pollen shed and access, will help to overcome these limitations

(Whitford et al., 2013). Hybrid breeding relies on the concept of heterosis, which leads to an increased vigor of an F1 genotype compared to its two parents. To maximize the heterosis effect, the parents that are crossed need to be fully homozygous and originate from genetically divergent heterotic groups (Melchinger, 1999). In wheat, however, the intensive germplasm exchange between breeding programs have greatly reduced the differentiation between gene pools to an extent that there are no currently available heterotic groups that can be used to maximize heterosis in hybrid breeding approaches (Longin et al., 2012). One potential strategy could be to cross lines from dissimilar target environments in order to expand the genetic diversity among pools (Whitford et al., 2013). Another approach proposed by Zhao et al. (2015) aims at the identification of heterotic patterns in elite germplasm pools, using high-throughput genotype data from the 90K Illumina wheat SNP array in order to find optimal parental combinations for crossings that maximize the hybrid vigor in the resulting F1 populations.

On a global scale, yield stability plays a major role in the face of unprecedented climatic fluctuations and the increasing frequency of severe drought and extreme heat events (Lesk et al., 2016). Increasing yields in marginal growing regions with very low regional production volumes, by generating more efficient cropping systems with a combination of better adapted, drought-tolerant varieties and improved farming practices, is of particular importance to meet the drastically increasing future demand for wheat (Tester and Langridge, 2010). Dissecting the complex genetic basis behind tolerance to water-limited environments, however, is particularly difficult, because under water shortage several abiotic stresses can challenge the plants simultaneously, confounding the direct effects of water limitation (Fleury et al., 2010). Effective mechanisms have been identified that enable some plant genotypes to regulate water supply and demand (Borrell et al., 2014). Tester and Langridge (2010) consider it possible that transgenic

approaches could help to precisely modify genes which directly affect drought tolerance. One example is was demonstrated in maize, where increased expression of a nuclear factor Y gene in transgenic lines significantly improved yields under drought conditions (Nelson et al., 2007).

The need to advance wheat productivity under unfavorable conditions, by improving resilience to instable climatic conditions in the course of climate change, has increased attention on the roots as the primary interface for water and nutrient acquisition. Several wheat studies underline the vital role of plants' "hidden half", including their importance for nitrogen and phosphorus uptake (Waines and Ehdaie, 2007), the connection between root proliferation and water/nutrient uptake ability (Den Herder et al., 2010), and their positive influence on grain yield (Atkinson et al., 2015). Increasing the root-to-shoot biomass ratio by allocating more root biomass in the plant's early life cycle at comparatively low shoot vigour is assumed to provide an adaptive advantage towards pre-anthesis water shortage (Manschadi et al., 2008). Roots also assume an important function in the physical refinement of plant stability, which is critical in approaches that try to increase yields by increasing the number of spikelets per plant, for example (Ren et al., 2012). Paradoxically, however, some studies report a significant diminishment of overall root systems in modern high-yielding varieties, a phenomenon that is often attributed to strong selective breeding under optimal water supply and fertilization (Den Herder et al., 2010; Ehdaie et al., 2003). It is furthermore assumed that "unconscious selection" via linkage drag, the unintended introduction of genes that are linked to a target gene, has led to non-optimal root systems in several "Green Revolution" wheat cultivars. This is most likely associated with the fact that the three bread wheat parents of the green-revolution wheats have less than two thirds of root biomass compared to several landraces (Waines and Ehdaie, 2007). Compared to the multitude of research that investigated the genetic architecture of various above-ground plant traits, the number of

molecular studies trying to identify candidate genes that influence wheat root growth is negligible. The few available publications describe QTL mapping approaches using bi-parental mapping populations (Bai et al., 2013; Comas et al., 2013; Manschadi et al., 2008; Sharma et al., 2011). Conventional QTL mapping is, however, fundamentally limited by its ability to assay only allelic diversity that segregates between parental lines, and by the intrinsically low levels of recombination in bi-parental mapping populations. The latter results in a low resolution of the QTL localization (Korte and Farlow, 2013). In contrast, GWAS approaches overcome these approaches by making use of usually larger natural populations. Natural, non-related populations cover a much broader range of genetic diversity than bi-parental populations, and represent a vastly greater number of recombination events. Despite these great advantages, however, high-resolution GWAS approaches have not been applied in hexaploid wheat root genomic approaches, most likely because it can be very challenging to accurately phenotype root traits in large diversity collections. To date the only association mapping study of root architectural traits in wheat was performed in seedlings from a collection of 183 tetraploid wheat lines, using 957 DArT and SSR markers (Cane et al., 2014). Due to the notorious measuring difficulty of root traits and the costs involved, molecular studies that supply reliable diagnostic markers for genomics-based selection approaches in modern breeding programs are strongly required (Ren et al., 2012).

Today, there are also public concerns that strong selective breeding has narrowed down the germplasm base to an extent that breeders cannot create sufficient initial variation for selection in their breeding programs, resulting in limitations for breeding progress (Reif et al., 2005). One approach that has been postulated for the genetic enrichment of elite breeding pools is to exploit crop wild relatives by introducing novel genetic diversity from ancient related species (Longin

and Reif, 2014). More than 560,000 wheat accessions are stored in international gene banks, representing a rich genetic resource to reinstate variation of genetic bottlenecks (e.g. from domestication or selective breeding). Amongst these accessions, many are already adapted to very specific target environments, possessing exclusive kinds of advantageous characteristics such as resistances towards specific biotic and abiotic stresses (Huang and Han, 2014; Lopes et al., 2015). In practice, however, introgressing diversity from very distant gene pools into elite wheat lines is often not trivial, due to poor arable performance of exotic lines in target environments for modern agriculture. Using high-resolution genome-wide marker information can help to effectively track favorable and unfavorable genetic variants in introgression lines, facilitating the required selection and backcrossing steps. Ancient lines are furthermore an excellent genetic resource for the establishment of large structured populations, such as nested association mapping (NAM) populations that improve the statistical power and resolution in genome-wide association mapping approaches, for the identification of candidate genes with an effect on important agronomic traits (Snowdon et al., 2015; Yu et al., 2008). The initial step for the exploitation of this untapped treasure of diversity is a large-scale genotyping of diverse germplasm (Massawe et al., 2016). With constantly improving ultrahigh-throughput genotyping tools, genomic characterization of large populations with many thousands of accessions is already feasible. This information, in combination with comprehensive data from advancing high-throughput phenotyping platforms, including remote sensing, field-based robotics and geo-referenced areal image capture with multisensoric imaging systems (Fahlgren et al., 2015), will help to estimate the “true breeding value” of these diverse lines and will sustain a targeted choice of accessions in the establishment of pre-breeding populations in breeding programs (Longin and Reif, 2014).

1.5 Scope and aims

The global demand for wheat is expected to strongly grow in the next three decades. However, worldwide trends show that yield increases in major growing regions tend to have reached a plateau. This has mainly been attributed to the low level of genetic diversity within elite wheat germplasm pools. Reintroducing novel genetic diversity from genetically distant material is seen as a key strategy to enhance the overall diversity, as a basis for breeders to reinvigorate breeding progress. For this purpose, deep knowledge about population stratification, genetic diversity and structural genomic variation is fundamentally important. With the advent of next-generation DNA sequencing technologies, high-throughput genotyping tools have evolved for all important crop plants, acquiring large-scale, high-resolution genomic insights that can be used to identify novel target genes that affect agronomic traits, to predict breeding values of untested lines, or for deep characterization of breeding material (Chapter 2). The work described in this dissertation aimed to analyze population structure, genetic diversity, linkage disequilibrium (LD) and hotspots for directional selection in a diverse collection of 460 hexaploid wheat lines, using a high-density genotyping array carrying 90,000 SNP markers (Chapter 3). The resulting data and insights provide breeders a means to select suitable crossing parents for precise reconstitution of genetic variation in genetically depleted genomic regions, enhancing rejuvenation of diversity controlling agronomically important traits.

To secure future wheat production in the face of climate change, particularly in marginal growing regions, the resilience of modern cultivars to various kinds of biotic and abiotic stresses is of paramount importance. This has increased attention on roots as the primary interface for nutrient and water acquisition. While general trends show that wheat root systems of modern varieties have been continuously diminished in the course of selective breeding, knowledge about the

genetic factors that control below-ground development of wheat is very limited. This dissertation describes the first genome-wide association study (GWAS) to date describing genomic regions associated with overall root proliferation in hexaploid wheat. The results suggest that strong selection for spike development has inadvertently reduced root variation due to linkage drag in modern elite lines (Chapter 4). These findings will help breeders to reverse this inadvertent co-selection. Simultaneously, candidate genes were identified that show expression in both root and spike tissues, and are also highly conserved across rice and sorghum, making them ideal targets for ongoing functional validations to decipher the genetics of underground plant architecture in important global cereal crops.

2 Understanding and utilizing crop genome diversity via high-resolution genotyping

Voss-Fels, K. and Snowdon, R. J

Plant Biotechnology Journal (2016) 14, 1086–1094

doi: 10.1111/pbi.12456

Plant Biotechnology Journal

Plant Biotechnology Journal (2016) **14**, pp. 1086–1094 doi: 10.1111/pbi.12456

Review article

Understanding and utilizing crop genome diversity via high-resolution genotyping

Kai Voss-Fels and Rod J. Snowdon*

Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany

Received 2 June 2015;
revised 17 July 2015;
accepted 20 July 2015.
**Correspondence* (Tel +49 641 9937420;
fax +49 641 9937429;
email rod.snowdon@agrar.uni-giessen.de)

Keywords: genome diversity, high-throughput genomics, selection, breeding.

Summary

High-resolution genome analysis technologies provide an unprecedented level of insight into structural diversity across crop genomes. Low-cost discovery of sequence variation has become accessible for all crops since the development of next-generation DNA sequencing technologies, using diverse methods ranging from genome-scale resequencing or skim sequencing, reduced-representation genotyping-by-sequencing, transcriptome sequencing or sequence capture approaches. High-density, high-throughput genotyping arrays generated using the resulting sequence data are today available for the assessment of genomewide single nucleotide polymorphisms in all major crop species. Besides their application in genetic mapping or genomewide association studies for dissection of complex agronomic traits, high-density genotyping arrays are highly suitable for genomic selection strategies. They also enable description of crop diversity at an unprecedented chromosome-scale resolution. Application of population genetics parameters to genomewide diversity data sets enables dissection of linkage disequilibrium to characterize loci underlying selective sweeps. High-throughput genotyping platforms simultaneously open the way for targeted diversity enrichment, allowing rejuvenation of low-diversity chromosome regions in strongly selected breeding pools to potentially reverse the influence of linkage drag. Numerous recent examples are presented which demonstrate the power of next-generation genomics for high-resolution analysis of crop diversity on a subgenomic and chromosomal scale. Such studies give deep insight into the history of crop evolution and selection, while simultaneously identifying novel diversity to improve yield and heterosis.

Genotyping-by-sequencing in crop plants for discovery of DNA sequence diversity

Novel genomic technologies have achieved exceeding importance for modern crop improvement and are undergoing continual further development in terms of efficiency and costs (Poland and Rife, 2012). Fifteen years after the complete genome sequence of the model plant *Arabidopsis thaliana* was decoded (genome size: 125 mega base pairs) (The Arabidopsis Genome Initiative, 2000), followed shortly afterwards by rice (430 Mbp) (Goff *et al.*, 2002), the rapid advancement of next-generation sequencing (NGS) platforms has today provided reference sequences for the large, complex genomes of many important crop species, such as maize (2500 Mbp) (Schnable *et al.*, 2009), sorghum (730 Mbp) (Paterson *et al.*, 2009), soya bean (1115 Mbp) (Schmutz *et al.*, 2010), potato (850 Mbp) (Xu *et al.*, 2011), barley (5100 Mbp) (Mayer *et al.*, 2012) and rapeseed (1200 Mbp) (Chalhoub *et al.*, 2014). Even in the huge hexaploid genome of bread wheat (17 000 Mbp), a combination of flow cytometry and syntenic mapping with next-generation sequencing technologies has enabled generation of a chromosome-based draft genome sequence (International Wheat Genome Sequencing Consortium, 2014). Pan-genome diversity analysis based on assemblies of crop relatives provides unprecedented insight into the gene diversity available in secondary crop gene pools. Recent pan-genome sequencing studies in maize (Lu *et al.*, 2015) and soya bean (Li *et al.*, 2014) describe the valuable contribution to pan-genomic variation to phenotypic variation for important adaptive traits. Identification and implementation of such adaptive potential using high-resolution genotyping may be a key to targeted rejuvenation of depleted phenotypic diversity in response to climate change, for example.

High-resolution genome information is being increasingly used by plant breeders to characterize germplasm, to identify genes that underlie important agronomic traits or to estimate the breeding values of individuals in breeding programmes in order to accelerate the selection of improved varieties (Varshney *et al.*, 2014). Either with or without a completed reference sequence, the development of reduced-representation genotyping-by-sequencing (GBS) methods has opened the way to use NGS technologies for high-throughput genomic resequencing, even in large plant populations, at a constantly shrinking price (Deschamps *et al.*, 2012). In maize, for example, Romay *et al.* (2013) used the GBS approach described by Elshire *et al.* (2011) to generate almost 700 000 genomewide SNPs in a panel of 2815 diverse inbred lines from globally distributed breeding programmes. Comprehensive sequencing data sets of this kind enable extremely high-resolution evaluation of genetic diversity and population structure, providing insight into the history of recombination and allelic diversity throughout different breeding pools (Qian *et al.*, 2014; Voss-Fels *et al.*, 2015). The broad applicability of GBS techniques for genetic analysis has been

successfully demonstrated in numerous important crops, for example rice (Spindel *et al.*, 2013), barley (Elshire *et al.*, 2011), potato (Uitdewilligen *et al.*, 2013), wheat (Poland *et al.*, 2015) and soya bean (Jarquín *et al.*, 2014). In combination with quantitative phenotype analysis in segregating populations, NGS methods also provide a powerful basis for rapid mapping and identification of genes underlying quantitative traits (e.g. Abe *et al.*, 2012; Gao *et al.*, 2013; Schneeberger *et al.*, 2009).

Targeting of genic variants associated with agronomic traits

Reduced-representation sequencing approaches involving exome capture or transcriptome sequencing enable targeted identification of molecular variants in protein-coding genome regions (Ku *et al.*, 2012). Reference-based assembly of target-captured or transcriptome sequence data in a test panel can allow rapid discovery of hundreds of thousands of molecular variants in gene-coding regions. High-density genic polymorphism data generated via such techniques can be applied to quantitative trait dissection, marker-assisted breeding, genomic selection or for high-resolution exploration of genetic resources (Bolger *et al.*, 2014). Custom design of capture probes or tailed sequencing primers provides a flexibility to target specific chromosome regions harbouring quantitative trait loci, known pathways from related model plants, or candidate genes for traits of interest across a given species of interest. For example, Gholami *et al.* (2012) and Rife *et al.* (2015) describe how tailed PCR can be used to target specific genes for next-generation resequencing in large numbers of individuals, while Schiessl *et al.* (2014) demonstrate the use of bead-based capture technology to resequence a panel of over 30 genes involved in regulating the flowering time pathway. Clarke *et al.* (2013) present an example for resequencing of genetic diversity spanning important meta-QTL in a major crop using a microarray-based capture platform. Harper *et al.* (2012) introduced the concept of associative transcriptomics, in which polymorphic SNP data from transcriptome sequencing in a diversity panel are associated with phenotype variation for QTL identification, and Mascher *et al.* (2013) describe how exome capture sequencing can help in the cost-effective identification of coding sequence variants even in very large genomes like that of barley. Collectively, these methods proved powerful options for mapping and discovery of genes underlying quantitative traits, and development of tightly linked, sequence-based markers for breeding. They also demonstrate the broad diversity of available technological platforms for sequence capture, which enable extremely flexible scaling of resequencing experiments to deal with few to many genes at low cost in large test populations.

Accessing crop diversity with high-density genotyping arrays

Beyond their direct applications for genetic mapping, QTL dissection and characterization of diversity, NGS technologies have also created the basis for the development of high-density SNP genotyping platforms as a high-throughput tool for genetic analysis of large experimental and breeding populations (Ganal *et al.*, 2012). Today, high-capacity SNP arrays are available for a broad range of plant species and are becoming widely used in breeding of major crops like maize (50–600 k SNPs) (Ganal *et al.*, 2011; Unterseer *et al.*, 2014), rice (51.5 k SNPs) (Chen *et al.*, 2014; Zhao *et al.*, 2011), wheat (9 k, 35 k, 90 k and 800 k SNPs) (Cavanagh *et al.*, 2013; Wang *et al.*, 2014; M. Winfield, A. Allen, A. Burridge, G. Barker, H. Benbow, P. Wilkinson, J. Coghill, C. Waterfall, A. Devassi, G. Scopes, T. Webster, F. Brew, C. Bloor, J. King, S. Griffiths, I. King, A. Bentley and K. Edwards *et al.*, unpublished), potato (8.3 k SNPs) (Hamilton *et al.*, 2011), barley (9 k SNPs) (Comadran *et al.*, 2012), soya bean (50 k SNPs) (Song *et al.*, 2013), rapeseed (60 k SNPs), (Edwards *et al.*, 2013) or sorghum (3 k and 90 k SNPs) (Bekele *et al.*, 2013; Wieckhorst *et al.*, 2015). At present, the Infinium® platform from Illumina Inc. (San Diego, CA) and the Axiom® technology from Affymetrix Inc. (Santa Clara, CA) are the most widely used platforms for large-scale SNP genotyping in crop plants (Thomson, 2014). Fixed genotyping chips are often preferred to GBS technologies for scenarios aiming to generate structured data sets of common sequence variants at low cost, with minimal bioinformatic input, for example within an ongoing breeding programme (Varshney *et al.*, 2014). On the other hand, to be effective a fixed SNP genotyping platform must be applicable to a wide range of different genotypes; hence, the alleles of the chosen SNPs must be representative even for diverse germplasm. GBS-based genotyping methods can be more suitable for identifying true, causal genetic variants for phenotypes with a complex genetic architecture, as these are typically influenced in crop species by rare alleles that may not be adequately represented on a SNP array (Huang and Han, 2014). Nevertheless, given the relatively high extent of linkage disequilibrium (LD) throughout the genomes of most crops, SNP markers on a fixed, high-density array are still likely to exhibit genetic associations with phenotypic variation through LD to the causal genes (Wray *et al.*, 2013).

Genome-scale characterization of crop diversity

High-throughput genotyping techniques are an important enabling technology for complex trait dissection by genomewide association studies (GWAS). Detailed molecular characterization of breeding germplasm, providing comprehensive knowledge of population genetic parameters and their relationships to natural and artificial selection for important traits, is a crucial prerequisite for the production of new, improved cultivars. Due to intensive human-mediated selection during plant breeding, modern crop varieties typically exhibit lower levels of genetic variation and biased allele frequency spectra compared to their wild types, especially in chromosome regions that harbour agronomically important loci (Mace *et al.*, 2013). This causes higher levels of overall, chromosome-wise and/or region-specific LD in the respective genomic regions. Agricultural selection furthermore has led to enlarged haplotypes with extended homozygosity, sometimes covering large chromosome segments (Mace *et al.*, 2013; Qian *et al.*, 2014; Voss-Fels *et al.*, 2015).

The enrichment of particular allele variants in gene pools due to directional selection, and the consequential depletion of genetic variation, caused genetic bottlenecks during crop domestication that have resulted in prominent selective sweeps in all major crops (Shi and Lai, 2015). This is particularly troublesome because of linkage drag, the unintentional co-selection of undesirable gene variants that are closely linked to selected loci of interest (Langridge and Fleury, 2011). Because genetic diversity represents the fundamental key to breeding success and a broad variation provides the basis for breeders to select varieties with constantly improving yield performance, these footprints of directional

selection seriously challenge crop improvement, as they lead to deterioration of genetically determined phenotypic variation. High-throughput population genomic analyses can address this dilemma by providing a detailed molecular basis for identification and genomic introgression of novel variation into chromosome segments surrounding directionally selected loci. Using high-resolution tools, we are now able to identify and characterize genome regions in most need of rejuvenation with novel diversity, and utilize genomic selection approaches (Jannink *et al.*, 2010) for the enrichment of depleted gene pools.

Identifying and overcoming signatures of selection

Jiao *et al.* (2012) used genome sequences of 278 maize inbred lines from China and the United States to describe the structural development of the maize genome during domestication and utilization by humans, identifying highly dynamic genetic changes caused by modern breeding. Based on around 4.8 million SNPs present in at least 50% of the population, they identified around 400 different loci that have experienced a selective sweep in different germplasm groups, representing domestication and modern breeding progress. They showed that modern breeding caused an accumulation of rare alleles in elite cultivars, suggesting that the proportion of rare alleles could be used as a selection index in future maize-breeding approaches.

A high-resolution investigation of genome-scale diversity and directional selection in sorghum by Mace *et al.* (2013) used the genome sequences of 44 highly diverse accessions, representative of the diversity spanning the primary gene pool. Strong signatures of selection were identified in different gene pools around major genes related to domestication, eco-geographical adaptation or agricultural traits such as plant height, seed colour and maturity. Investigation of genomewide diversity patterns revealed decreased diversity in landraces and improved germplasm, but discovered untapped genetic variation in related allopatric gene pools (Figure 1; Mace *et al.*, 2013). Similarly, in rice, Huang *et al.* (2012) resequenced 446 diverse accessions of the rice wild relative *Oryza rufipogon*, along with 1083 cultivated *O. indica* and *O. japonica* varieties. This study revealed 55 selective sweeps originating from domestication. Around 8 million SNP markers disclosed extremely high allelic variation in wild rice populations compared to domesticated rice, highlighting genomic target regions for the restoration of genetic diversity in future rice-breeding efforts.

In many crops, the co-selection of undesired loci due to linkage drag has hampered the efficiency of introgression approaches using exotic plant resources, and the resolution with which introgressed DNA segments could be tracked on a molecular level was extremely poor when concepts for marker-assisted backcrossing were first introduced to breeding. Massively parallel genotyping techniques overcome this dilemma and facilitate marker-assisted selection in very early stages of plant development and the breeding cycle. In a simulation study using maize data, Herzog *et al.* (2014) demonstrated the inherent suitability of high-throughput genotyping arrays for targeted introgression of donor chromosome segments into recipient genotypes. Detection of introgressed DNA fragments using detailed molecular marker information facilitates the utilization of exotic germplasm for the targeted restoration of genetic diversity in crop breeding populations, improving our capacity to introduce novel loci into elite cultivars with minimal linkage drag.

Even in complex polyploid genomes, the use of large-scale molecular information from genomewide SNP markers can be used to reveal the comparative influence of artificial selection and breeding for important agronomic traits on LD and haplotype structure. Divergent bread wheat gene pools show extreme differences in local LD surrounding loci involved in important adaptation and grain quality traits, for example (Figure 2; Voss-Fels *et al.*, 2015). In the highly duplicated A and C subgenomes of rapeseed, Qian *et al.* (2014) demonstrated strong subgenomic bias for selection signatures during breeding for important seed quality traits. Serious erosion of genetic variability in C-subgenome QTL was found to reflect a considerably lower diversity and a reduced recombination rate, which in turn hamper breeding progress and heterotic potential. Such information provides breeders with important information to develop strategies to precisely reinstate diversity in such regions, for example by inducing elevated recombination via de novo polyploidization (Snowdon *et al.*, 2015).

Enriching subgenomic diversity for improvement of heterotic potential

Translating high-resolution genome data to structured breeding populations derived from sequenced founder lines provides a basis for knowledge-based enrichment of low-diversity chromosome segments, to overcome linkage drag associated with large selection signatures (Voss-Fels *et al.*, 2015). This can also enhance heterosis by expanding diversity between hybrid breeding pools into chromosome segments containing strong adaptation signatures. Snowdon *et al.* (2015) present the concept of heterotic haplotype capture (HHC), which uses whole-genome profiling to identify and enrich diversity-poor genome regions and introduce these into hybrid breeding programmes for targeted improvement of heterosis. In HHC, fully sequenced, diverse founder lines are used to generate large structured prebreeding populations, like nested association mapping panels (Gore *et al.*, 2009; Yu *et al.*, 2008). By genotyping an entire population with a high-density SNP array, it is possible to detect crossover breakpoints in every individual at a previously unavailable resolution and combine these with parental sequence data to impute sequenced haplotypes across the whole genome in huge populations. Sequencing of male-sterile maternal lines, used to create and phenotype test hybrids from such a population, can subsequently provide a basis to identify haplotype structures associated with heterotic trait performance (Snowdon *et al.*, 2015). The HHC concept thus enables introduction and characterization of novel diversity on a high-resolution, subchromosomal level, while simultaneously facilitating the replenishment of eroded diversity in strongly selected genome regions (Figure 3).

Unravelling the genetic basis of heterosis

Besides the high-definition molecular characterization and utilization of crop breeding germplasm, novel genotyping technologies can also shed new light on the genetic background of heterosis. Hybrid crops, exploiting heterosis for yield gain and stability, have become one of the major drivers of increased agricultural production during the past few decades. Despite the global importance of heterosis for food security, and the growing tendency towards utilization of hybrid vigour even in inbreeding crops like bread wheat, the molecular and genetic mechanisms underlying this phenomenon are still not completely understood.

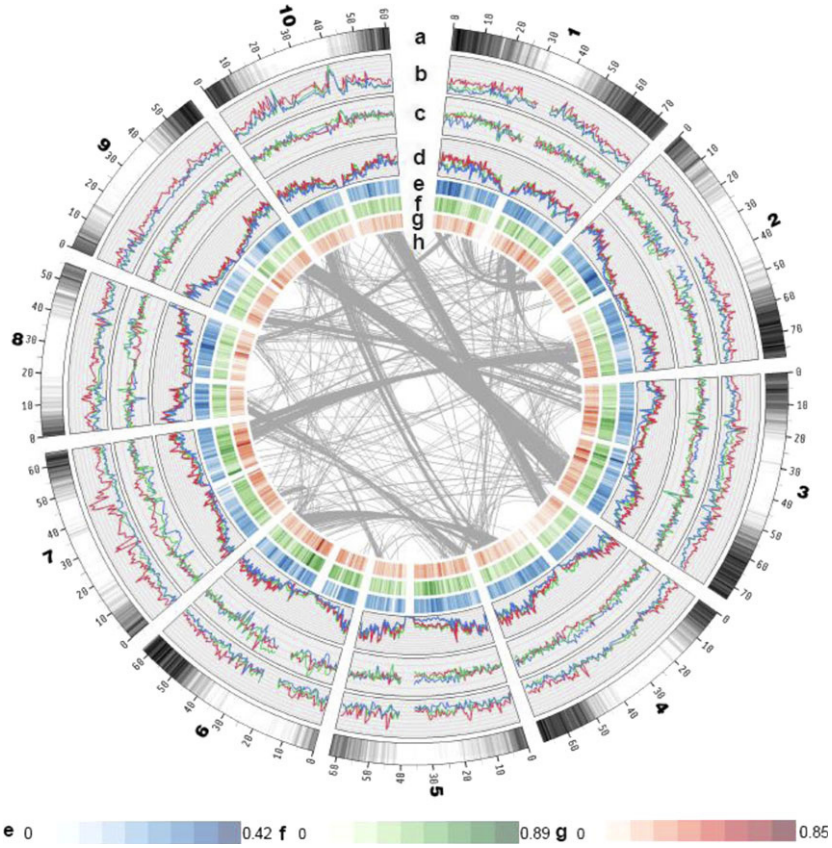


Figure 1 Genomewide patterns of sequence diversity in *Sorghum bicolor*. The 10 chromosomes are portrayed along the perimeter of each circle. Concentric circles display (a) gene content density distribution; (b) genomic diversity of wild and weedy genotypes (red), landraces (green) and improved inbreds (blue); (c) selection patterns (Tajima's D statistic) in wild and weedy genotypes (red), landraces (green) and improved inbreds (blue); (d) number of SNPs in wild and weedy genotypes (red), landraces (green) and improved inbreds (blue); (e) fixation indices (F_{ST} values) of improved inbreds versus landraces; (f) F_{ST} values of improved inbreds versus wild and weedy genotypes; and (g) F_{ST} values of landraces versus wild and weedy genotypes. (h) A graphical view of duplicated annotated genes is indicated by connections between segments. Figure from Mace *et al.* (2013), reprinted with permission from Macmillan Publishers Ltd, *Nature Communications* © 2013.

Genome profiling in large hybrid populations today offers an unprecedented resolution for dissection of loci and genes involved in heterotic expression. Recently, Huang *et al.* (2015) described a study in which an extensive population of 1495 elite hybrid rice varieties, along with their inbred parental lines, was subject to detailed genomewide sequence analysis in order to investigate genomic effects on hybrid vigour for 38 agronomic traits. The resequenced genomes of all parental lines harboured around 1.3 million polymorphic SNP markers, which were subsequently used to study population genetic parameters and perform GWAS at an unprecedented resolution. In particular, heterozygous chromosome regions were revealed that contribute to trait expression in the F_1 hybrids. Elucidation of the corresponding genomic effects on phenotypic traits demonstrated that pyramiding of multiple loci facilitates the accumulation of numerous rare superior alleles with positive effects. In other words, dominance complementation contributes most to the heterosis effect in hybrid rice production. A combination of forward and background selection using high-throughput genome screening tools (Herzog and Frisch, 2011; Herzog *et al.*, 2014) can thus be expected to significantly improve potential for increasing breeding gain through efficient exploitation of hybrid vigour.

The idea of genomic hybrid breeding, in which a genome-based prediction strategy based on genome sequence data is applied to estimate the performance of the F_1 progeny in hybrid breeding, was introduced in rice by Xu *et al.* (2014). Using over 250 000 SNP markers, generated by resequencing 210 parental inbred lines from a training set of 278 randomly selected hybrids, this study demonstrated the power of marker-directed estimation of F_1 hybrid yields in rice. The top 100 predicted hybrids, from a total of 21 945 possible combinations between the parental accessions, were estimated to exceed the overall yield average by 16%. This represents a significant improvement on average selection gains from conventional breeding and accelerated hybrid rice production.

Further applications for breeding and crop improvement

High-throughput, high-density genome-profiling tools enable the rapid and low-cost portrayal of crop genome characteristics in a precise, high-resolution manner. Identification of molecular variants on the DNA sequence level opens versatile options for practical application. As described above, breeders can use this detailed information for more targeted germplasm interchange between gene pools for improvement of diversity and heterosis. Generally, crop wild relatives represent an excellent genetic resource to reinstate variation caused by genetic bottlenecks during crop domestication and breeding (Huang and Han, 2014; Lopes *et al.*, 2015). On the other hand, breeders are often reluctant to implement completely novel diversity from distant gene pools due to their lack of adaptation and the consequent performance penalty. High-density genome data may help to improve this problem by enabling more efficient genomic selection strategies that efficiently predict performance based on genomewide marker combinations (Heffner *et al.*, 2009).

One key to improving crop performance and breeding processes is the enhancement of recombination in diversity-poor chromosome regions. Two recent studies have demonstrated how this might be achieved with support from high-resolution genotype data. By high-coverage sequencing of 41 rice offspring

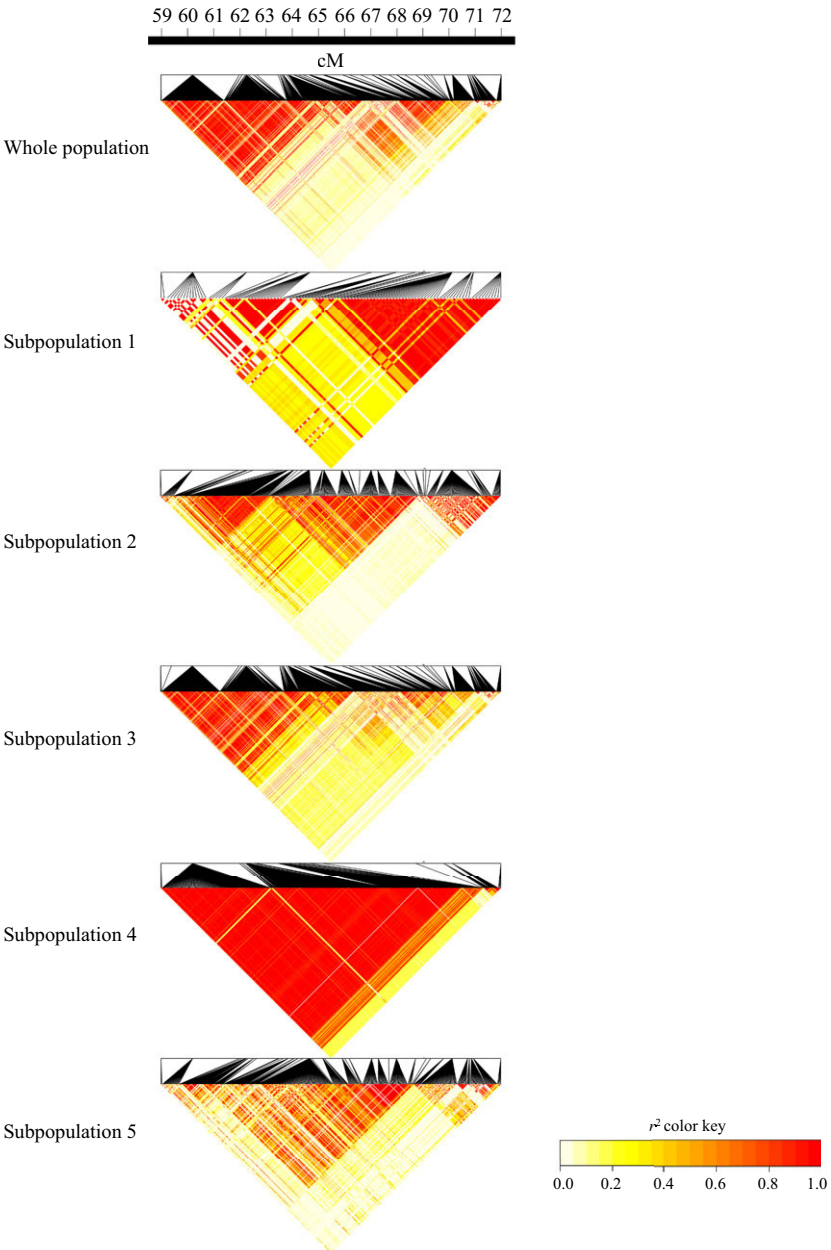


Figure 2 Detailed comparison of local linkage disequilibrium (LD) decay on a 13-cM segment of *Triticum aestivum* chromosome 1B in a population of 460 international wheat accessions, comparing the local genetic diversity within five subpopulations representing distinct breeding pools. This example demonstrates how strong directional selection in distinct breeding pools can lead to highly distinct patterns of LD. Densely spaced SNP markers can assist backcrossing programmes to enrich diversity-poor regions. Figure adapted from Voss-Fels *et al.* (2015), reprinted with permission from the Crop Science Society of America, *The Plant Genome* © 2015.

form a biparental cross, Si *et al.* (2015) generated a detailed map of recombination hot spots and cold spots based on 900 000 high-quality polymorphic loci. Interestingly, the recombination hot spot regions were enriched with genes involved in response to environmental stimuli, and environmental stress was found to increase the recombination rate in around one-third of the genotypes. If this can be confirmed, then breeders may be able to increase recombination by making crosses on plants grown under stressed conditions. In another approach, Suay *et al.* (2014) demonstrated that interspecific *Brassica* hybrids carrying a specific chromosome have a significantly elevated recombination rate and reduced crossover interference. As described by Mason *et al.* (2014), high-throughput genomewide SNP genotyping provides an ideal basis for molecular cytogenetic analysis of such materials to infer chromosome behaviour at meiosis and identify individuals with elevated recombination in diversity-poor chromosome regions for use in breeding.

Genomic selection models predict the breeding value of an individual based on molecular marker information, using statistical calibrations from representative test populations with robust phenotype data and genomewide SNP profiles. Genomic selection is becoming a widespread technique in breeding of many important crops, for example maize, wheat, rice and barley (Lin *et al.*, 2014; Poland *et al.*, 2012; Riedelsheimer *et al.*, 2012).

Despite the exceptional recent advances in molecular tools for genotyping, efficient, high-throughput phenotyping platforms are still a major bottleneck in the dissection and understanding of high-value, quantitatively inherited traits. Particularly for traits that are unable to be effectively assessed under controlled conditions, there is still a need for further improvements in automated phenotyping in field trials. Recent advances in remote sensing, field-based robotics and geo-referenced aerial image capture with multisensoric imaging systems (Fahlgren *et al.*, 2015), in combination with high-performance computing, are

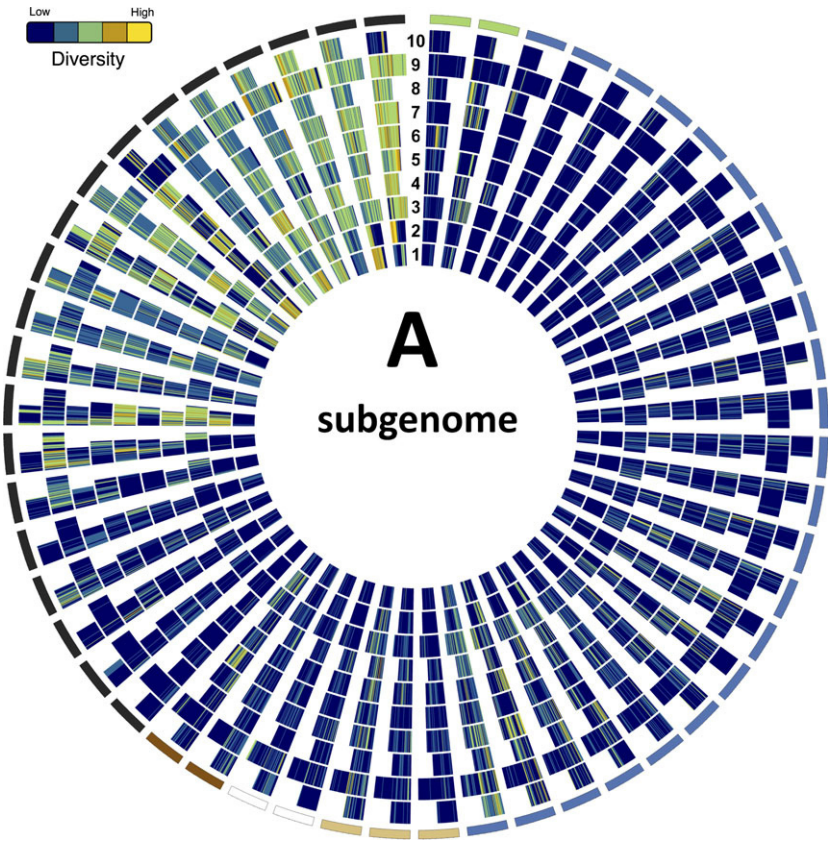


Figure 3 Genomics-assisted rejuvenation of a depleted breeding pool. The founder lines from the *Brassica napus* nested association mapping (NAM) panel (represented in this example by segments describing the 10 chromosomes of the *B. napus* A subgenome) introduce completely novel diversity into the depleted gene pool of cultivated winter oilseed rape. Diversity erosion (dark blue), displayed by many chromosomes of natural *B. napus* (coloured and white bars in external ring), is replaced by considerably enriched sequence diversity (yellow) and strong recombination in the corresponding chromosome regions of many synthetic forms (black bars in external ring). Figure adapted from Snowdon *et al.* (2015) and reprinted with permission from Elsevier, *Trends in Plant Science* © 2015.

improving the phenotyping bottleneck (Araus and Cairns, 2014; White *et al.*, 2012). Standardization of high-throughput phenotype ontologies for automated analysis in association with genotype data remains a challenge, and effective data management and interpretation pipelines are still required to increase the applicability of high-throughput phenotyping platforms in crop improvement.

Conclusions and outlook

The global demand for major crops is expected to strongly increase as a consequence of the steadily growing world population. On the other hand, the decreasing agricultural area, and increasingly stressed environments for plant production in the face of climate change, challenges plant breeders worldwide to produce constantly improving varieties that produce high and stable yields. The fundamental key to breeding success is genetic diversity, which provides breeders the basis for the selection of cultivars with a superior arable performance. The availability of powerful genomics tools provides an unprecedented basis to accelerate crop improvement and increase genetic gain in breeding programmes, especially for traits which are difficult, time-consuming or expensive to accurately evaluate phenotypically. The genome sequences of the world's most important crop plants open the way to identify and characterize all available diversity not only within crop primary gene pools, but also throughout related species. Ultra-high-throughput genotyping techniques like GBS or array-based SNP genotyping enable the prompt genome profiling of even large plant populations at constantly shrinking costs, providing breeders detailed, high-resolution molecular information. On the one hand, this can greatly improve the genetic resolution for mapping and cloning of useful genes and QTL, particularly as large structured populations like NAM or HHC panels become available for important crop species. High-resolution analysis of breeding germplasm allows breeders to gain deep and highly precise insights into genetic diversity on a subgenomic and chromosomal level. This is particularly helpful for targeted enrichment of depleted gene pools, in which strong, human-mediated selection during domestication and breeding has caused a dramatic loss of genetic diversity in many genome regions of modern varieties. Chromosome segments harbouring loci involved in essential traits such as vernalization requirement, winter hardiness, flowering time, seed quality or resistances against biotic stress are often associated with large blocks of very strong LD which can negatively impact yield via linkage drag. The ability to precisely identify these signatures of directional selection provides a basis to precisely reinstate diversity in affected genomic regions. Detailed genome profiles can also be very helpful to identify germplasm that is most suitable for crossing in order to reverse the erosion of genetic variation in elite material. Genomics-directed strategies to simultaneously increase recombination rates can facilitate this process. The targeted introgression of loci from exotic germplasm into high-performance varieties is likely to play a key role in continual improvement of heterotic performance in major crops. Ultimately, genomic selection and hybrid prediction strategies based on cheap, high-throughput, high-density genetic marker data will be a key factor in the acceleration of breeding progress to provide food, feed, fibre and fuel for future generations. Translating the huge progress in low-cost, high-resolution genotyping to significant increases in crop improvement requires similar advances in the development, cost and field applicability of high-throughput

phenotyping tools, including improved automated analysis and interpretation of geo-referenced, multisensor, field-based crop imaging data.

Acknowledgements

The authors were supported for this work by funding from the German Research Foundation (DFG grant number SN14/16-1), the Federal Ministry of Food and Agriculture (BMEL grant number FNR-22408212) and the Federal Ministry of Food and Agriculture (BMEL grant number 0315964H).

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S. and Terauchi, R. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotech.* **30**, 174–178.

Araus, J.L. and Cairns, J.E. (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* **19**, 52–61.

Bekele, W.A., Wieckhorst, S., Friedt, W. and Snowdon, R.J. (2013) High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array. *Plant Biotech. J.* **11**, 1112–1125.

Bolger, M.E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B. and Mayer, Klaus.F.X. (2014) Plant genome sequencing - applications for crop improvement. *Curr. Opin. Biotech.* **26**, 31–37.

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., da Silva, M.L., Bockelman, H., Talbert, L., Anderson, J.A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P.L., Dubcovsky, J., Morell, M.K., Sorrells, M.E., Hayden, M.J. and Akhunov, E. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Nat. Acad. Sci. USA*, **110**, 8057–8062.

Chalhoub, B., Denoeud, F., Liu, S., Parkin Isobel, A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.-C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.-H., Thi, V.H.D., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom Christine, H.D., Wang, X., Canaguier, A., Chauveau, A., Bérard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.-M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C. and Wincker, P. (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**, 950–953.

Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y., Tang, X., Zhou, F., Deng, X.W. and Zhang, Q. (2014) A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant*, **7**, 541–553.

Clarke, W.E., Federico, M.L., Gajardo, H.A., Gerhardt, D.J., Higgins, E., Sharpe, A.G., Snowdon, R.J., Parkin, I.A.P. and Iniguez-Luy, F.L. (2013) Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. *PLoS ONE*, **8**, e81992.

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

Deschamps, S., Llaca, V. and May, G.D. (2012) Genotyping-by-sequencing in plants. *Biology*, **1**, 460–483.

Edwards, D., Batley, J. and Snowdon, R.J. (2013) Accessing complex crop genomes with next-generation sequencing. *Theor Appl Genet.* **126**, 1–11.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Fahlgren, N., Gehan, M.A. and Baxter, I. (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* **24**, 93–99.

Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.-M., Hansen, M., Joets, J., Le Paslier, M.-C., McMullen, M.D., Montalent, P., Rose, M., Schön, C.-C., Sun, Q., Walter, H., Martin, O.C. and Falque, M. (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE*, **6**, e28334.

Ganal, M.W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H. and Durstewitz, G. (2012) Large SNP arrays for genotyping in crop plants. *J. Biosci.* **37**, 821–828.

Gao, Z.Y., Zhao, S.C., He, W.M., Guo, L.B., Peng, Y.L., Wang, J.J., Guo, X.S., Zhang, X.M., Rao, Y.C., Zhang, C., Rao, Y.C., Zhang, C., Dong, G.J., Zheng, F.Y., Lu, C.X., Hu, J., Zhou, Q., Liu, H.J., Wu, H.Y., Xu, J., Ni, P.X., Zeng, D.L., Liu, D.H., Tian, P., Gong, L.H., Ye, C., Zhang, G.H., Wang, J., Tian, F.K., Xue, D.W., Liao, Y., Zhu, L., Chen, M.S., Li, J.Y., Cheng, S.H., Zhang, G.Y., Wang, J. and Qian, Q. (2013) Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. *Proc. Natl Acad. Sci. USA*, **110**, 14492–14497.

Gholami, M., Bekele, W.A., Schondelmaier, J. and Snowdon, R.J. (2012) A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants. *Plant Biotech. J.* **10**, 635–645.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.

Gore, M.Q., Chia, J.M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.L., McMullen, M.D., Grills, G.S., Ross-Ibarra, J., Ware, D.H. and Buckler, E.S. (2009) A first-generation haplotype map of maize. *Science*, **326**, 1115–1117.

Hamilton, J.P., Hansey, C.N., Whitty, B.R., Stoffel, K., Massa, A.N., van Deynze, A., Jong, D., Walter, S., Douches, D.S. and Buell, C.R. (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genom.* **12**, 302.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P. and Bancroft, I. (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotech.* **30**, 798–802.

Heffner, E.L., Sorrells, M.E. and Jannink, J.L. (2009) Genomic selection for crop improvement. *Crop Sci.* **49**, 1–12.

Herzog, E. and Frisch, M. (2011) Selection strategies for marker-assisted backcrossing with high-throughput marker systems. *Theor. Appl. Genet.* **123**, 251–260.

Herzog, E., Falke, K.C., Presterl, T., Scheuermann, D., Ouzunova, M. and Frisch, M. (2014) Selection strategies for the development of maize introgression populations. *PLoS ONE*, **9**, e92429.

Huang, X. and Han, B. (2014) Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **65**, 531–551.

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., Feng, L., Furuumi, H., Kubo, T., Miyabayashi, T., Yuan, X., Xu, Q., Dong, G., Zhan, Q., Li, C., Fujiyama, A., Toyoda, A., Lu, T., Feng, Q., Qian, Q., Li, J. and Han, B. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497–501.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., Chen, N., Hao, Z., Liu, K., Zhu, C., Huang, T., Zhao, Q.,

Zhang, L., Fan, D., Zhou, C., Lu, Y., Weng, Q., Wang, Z.-X., Li, J. and Han, B. (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Comm.* **6**, 6258.

International Wheat Genome Sequencing Consortium. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.

Jannink, J.L., Lorenz, A.J. and Iwata, H. (2010) Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics*, **9**, 166–177.

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G. and Lorenz, A. (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genom.* **15**, 740.

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B., Liu, Z., Chen, J., Li, W., Zhang, M., Xie, S. and Lai, J. (2012) Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **44**, 812–815.

Ku, C.-S., Wu, M., Cooper, D.N., Naidoo, N., Pawitan, Y., Pang, B., Iacopetta, B. and Soong, R. (2012) Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev. Mol. Diagn.* **12**, 241–251.

Langridge, P. and Fleury, D. (2011) Making the most of 'omics' for crop breeding. *Trends Biotechnol.* **29**, 33–40.

Li, Y.H., Zhou, G., Ma J., J., Jin, L., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., Zhang, S., Zuo, Q., Shi, X., Li, Y., Zhang, W., Hu, Y., Kong, G., Hong, H., Tan, B., Song, J., Liu, Z., Wang, Y., Ruan, H., Yeung, C.K.L., Liu, J., Wang, H., Zhang, L., Guan, R., Wang, K., Li, W., Chen, S., Chang, R., Jiang, Z., Jackson, S.A., Li, R. and Qiu, L. (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotech.* **32**, 1045–1052.

Lin, Z., Hayes, B.J. and Daetwyler, H.D. (2014) Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* **65**, 1177.

Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K., Aktas, H., Ozer, E., Ozdemir, F., Manickavelu, A., Ban, T. and Vikram, P. (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **66**, 3477–3486. doi:10.1093/jxb/erv122.

Lu, F., Romay, M.C., Glaubitz, J.C., Bradbury, P.J., Elshire, R.J., Wang, T., Li, Y., Li, Y., Semagn, K., Zhang, X., Hernandez, A.G., Mikel, M.A., Soifer, I., Barad, O. and Buckler, E.S. (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat. Commun.* **6**, 6914.

Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentis, P.J., Bian, L., Campbell, B.C., Hu, W., Innes, D.J., Han, X., Cruickshank, A., Dai, C., Frère, C., Zhang, H., Hunt, C.H., Wang, X., Shatte, T., Wang, M., Su, Z., Li, J., Lin, X., Godwin, I.D., Jordan, D.R. and Wang, J. (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **4**, 2320.

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M., D'Ascenzo, M., Akhunov, E.D., Hedley, P.E., Gonzales, A.M., Morrell, P.L., Kilian, B., Blattner, F.R., Scholz, U., Mayer, K.F., Flavell, A.J., Muehlbauer, G.J., Waugh, R., Jeddeloh, J.A. and Stein, N. (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**, 494–505.

Mason, A.S., Batley, J., Bayer, P.E., Hayward, A., Cowling, W.A. and Nelson, M.N. (2014) High-resolution molecular karyotyping uncovers pairing between ancestrally related *Brassica* chromosomes. *New Phytol.* **202**, 964–974.

Mayer, K.F.X., Waugh, R., Brown, J.W.S., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J., Wise, R.P. and Stein, N. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Rahman, M., Ware, D., Westhoff, P., Mayer Klaus, F.X., Messing, J. and Rokhsar, D.S. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Poland, J. (2015) Breeding-assisted genomics. *Curr. Opin. Plant Biol.* **24**, 119–124.

Poland, J.A. and Rife, T.W. (2012) Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome*, **5**, 92.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M. and Jannink, J.L. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, **5**, 103–113.

Qian, L., Qian, W. and Snowdon, R.J. (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genom.* **15**, 1170.

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L. and Melchinger, A.E. (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* **44**, 217–220.

Rife, T.W., Wu, S., Bowden, R. and Poland, J.A. (2015) Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. *BMC Genom.* **16**, 169.

Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-Garcia, S.A., McMullen, M.D., Holland, J.B., Buckler, E.S. and Gardner, C.A. (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, R55.

Schiessl, S., Samans, B., Hüttel, B., Reinhard, R. and Snowdon, R.J. (2014) Capturing sequence variation among flowering-time regulatory gene homologues in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* **5**, 404.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. and Jackson, S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. and Wilson, R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Lehmann Nielsen, K., Jørgensen, J. and Weigel Andersen, S.U. (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Meth.* **6**, 550–551.

Shi, J. and Lai, J. (2015) Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* **24**, 47–53.

Si, W., Yuan, Y., Huang, J., Zhang, X., Zhang, Y., Zhang, Y., Tian, D., Wang, C., Yang, Y. and Yang, S. (2015) Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F_2 plants. *New Phytol.* **206**, 1491–1502.

Snowdon, R.J., Abbadi, A., Kox, T., Schmutzer, T. and Leckband, G. (2015) Heterotic haplotype capture: precision breeding for hybrid performance. *Trends Plant Sci.* **20**, 410–413.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE*, **8**, e54985.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N. and McCouch, S. (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* **126**, 2699–2716.

Suay, L., Zhang, D., Eber, F., Jouy, H., Lodé, M., Huteau, V., Coriton, O., Szadkowski, E., Leflon, M., Martin, O.C., Falque, M., Jenczewski, E., Paillard, S. and Chèvre, A.-M. (2014) Crossover rate between homologous chromosomes and interference are regulated by the addition of specific unpaired chromosomes in *Brassica*. *New Phytol.* **201**, 645–656.

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Thomson, M.J. (2014) High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotech.* **2**, 195–212.

Uitdewilligen, J.G.A.M.L., Wolters, A.A., D'hoop, B.B., Borm, T.J.A. and Visser, R.G.F. (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE*, **8**, e62355.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T.M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K.F. and Schön, C.-C. (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genom.* **15**, 823.

Varshney, R.K., Terauchi, R. and McCouch, S.R. (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* **12**, e1001883.

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. and Snowdon, R.J. (2015) Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome*, **8**, doi:10.3835/plantgenome2015.03.0013.

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

White, J.W., Andrade-Sanchez, P., Gore, M.A., Bronson, K.F., Coffelt, T.A., Conley, M.M., Feldmann, K.A., French, A.N., Heun, J.T., Hunsaker, D.J., Jenks, M.A., Kimball, B.A., Roth, R.L., Strand, R.J., Thorp, K.R., Wall, G.W. and Wang, G. (2012) Field-based phenomics for plant genetics research. *Field. Crop. Res.* **133**, 101–112.

Wieckhorst, S., Bekele, W., Kloiber-Maitz, M., Schulz-Streeck, T., Knaak, C., Ouzunova, M., Davassi, A. and Snowdon, R. (2015) *A high-density SNP genotyping array for genome-based breeding of energy sorghum for central Europe. Plant and Animal Genome XXIII Conference, San Diego, CA, USA, January 10-14*, 2015. https://pag.confex.com/pag/xxiii/webprogram/Paper15935.html

Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E. and Visscher, P.M. (2013) Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., Orjeda, G., Guzman, F., Torres, M., Lozano, R., Ponce, O., Martinez, D., De la Cruz, G., Chakrabarti, S.K., Patil, V.U., Skryabin, K.G., Kuznetsov, B.B., Ravin, N.V., Kolganova, T.V., Beletsky, A.V., Mardanov, A.V., Di Genova, A., Bolser, D.M., Martin David, M.A., Li, G., Yang, Y., Kuang, H., Hu, Q., Xiong, X., Bishop, G.J., Sagredo, B., Mejía, N., Zagorski, W., Gromadka, R., Gawor, J., Szczesny, P., Huang, S., Zhang, Z., Liang, C., He, J., Li, Y., He, Y., Xu, J., Zhang, Y., Xie, B., Du, Y., Qu, D., Bonierbale, M., Ghislain, M., Herrera, M.D.R., Giuliano, G., Pietrella, M., Perrotta, G., Facella, P., O'Brien, K., Feingold, S.E., Barreiro, L.E., Massa, G.A., Diambra, L., Whitty, B.R., Vaillancourt, B., Lin, H., Massa, A.N., Geoffroy, M., Lundback, S., DellaPenna, D., Buell, C.R., Sharma, S.K., Marshall, D.F., Waugh, R., Bryan, G.J., Destefanis, M., Nagy, I., Milbourne, D., Thomson, S.J., Fiers, M., Jacobs Jeanne, M.E., Nielsen, K.L., Sønderkær, M., Iovene, M., Torres, G.A., Jiang, J., Veilleux, R.E., Bachem Christian, W.B., de Boer, J., Borm, T., Kloosterman, B., van Eck, H., Datema, E., Hekkert, B.T., Goverse, A., van Ham, R.C.H.J. and Visser, R.G.F. (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195.

Xu, S., Zhu, D. and Zhang, Q. (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Nat. Acad. Sci. USA*, **111**, 12456–12461.

Yu, J., Holland, J.B., McMullen, M.D. and Buckler, E.S. (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics*, **178**, 539–551.

Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D. and McCouch, S.R. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2**, 467.

3 Subgenomic diversity patterns caused by directional selection in bread wheat gene pools

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. and Snowdon, R. J.

The Plant Genome (2015) 8, 0.

doi: 10.3835/plantgenome2015.03.0013

Published July 10, 2015

ORIGINAL RESEARCH

Subgenomic Diversity Patterns Caused by Directional Selection in Bread Wheat Gene Pools

Kai Voss-Fels, Matthias Frisch, Lunwen Qian, Stefan Kontowski, Wolfgang Friedt, Sven Gottwald, and Rod J. Snowdon*

Abstract

Genetic diversity represents the fundamental key to breeding success, providing the basis for breeders to select varieties with constantly improving yield performance. On the other hand, strong selection during domestication and breeding have eliminated considerable genetic diversity in the breeding pools of major crops, causing erosion of genetic potential for adaptation to emerging challenges like climate change. High-throughput genomic technologies can address this dilemma by providing detailed knowledge to characterize and replenish genetic diversity in breeding programs. In hexaploid bread wheat (*Triticum aestivum* L.), the staple food for 35% of the world's population, bottlenecks during allopolyploidisation followed by strong artificial selection have considerably narrowed diversity to the extent that yields in many regions appear to be unexpectedly stagnating. In this study, we used a 90,000 single nucleotide polymorphism (SNP) wheat genotyping array to assay high-frequency, polymorphic SNP markers in 460 accessions representing different phenological diversity groups from Asian, Australian, European, and North American bread wheat breeding materials. Detailed analysis of subgroup diversity at the chromosome and subgenome scale revealed highly distinct patterns of conserved linkage disequilibrium between different gene pools. The data enable identification of genome regions in most need of rejuvenation with novel diversity and provide a high-resolution molecular basis for genomic-assisted introgression of new variation into chromosome segments surrounding directionally selected metaloci conferring important adaptation and quality traits.

Published in The Plant Genome 8
doi: 10.3835/plantgenome2015.03.0013

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

STAGNATING YIELDS IN BREAD WHEAT over the past two decades in Europe, North America, Asia, and Australia (Ray et al., 2012; Brisson et al., 2010) have raised concerns that reduction of genetic diversity through intensive breeding has narrowed the genetic potential for yield gain (Reif et al., 2005). Genomics technologies can provide detailed knowledge to overcome these problems and to create, broaden, and maintain genetic variation for wheat improvement.

Knowledge on the relationships, population structure, and genetic makeup of plant populations is a crucial prerequisite for the optimal selection of crossing parents in breeding programs (Zhang et al., 2011). In European wheat, Nielsen et al. (2014) used 1849 polymorphic diversity array technology (DArT) markers to demonstrate a clear subdivision into two main subgroups based on the allelic status of the major dwarfing locus *Rht8*. Intensive

K. Voss-Fels, L. Qian, W. Friedt, S. Gottwald, and R.J. Snowdon, Dep. of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig Univ., Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany; M. Frisch, Institute for Agronomy and Plant Breeding II, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig Univ., Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany; S. Kontowski, W. von Borries-Eckendorf GmbH & Co. KG, Hovedisser Str. 92, 33818 Leopoldshöhe, Germany.
Received 10 Mar. 2015. Accepted 5 May 2015. *Corresponding author (rod.snowdon@agrar.uni-giessen.de).

Abbreviations: DArT, diversity array technology; F_{ST}, Wright's fixation index; GBS, genotyping-by-sequencing; GS, genomic selection; GWAS, genome-wide association study; IBS, identity-by-state; LD, linkage disequilibrium; LOESS, locally estimated scatterplot smoothing; MAF, minor allele frequency; MRD, modified Roger's distance; PCA, principal component analysis; PIC, polymorphism information content; QTL, quantitative trait loci; SHA3/CBRD, 'Shanghai-3/Catbird'; SNP, single nucleotide polymorphism; SP, subpopulation; STS, sequence tagged site; UPGMA, unweighted-pair-group method with arithmetic mean.

selection at this locus was found to have a major influence on allelic variation on chromosome 2D. This is a good example of how genome-wide molecular markers can be used as a basis to pinpoint deficits in genetic diversity among elite wheat breeding populations.

Understanding of population stratification and genetic relatedness also provides a basis to establish genetically divergent heterotic groups for maximization of heterosis in hybrid breeding approaches (Melchinger, 1999), which are expected to considerably support wheat improvement in the near future (Longin and Reif, 2014). Genetically diverse gene pools represent the basis of heterotic potential. However, the intensive exchange of elite varieties within wheat breeding programs, which traditionally have focused on inbreeding strategies rather than hybrid pool formation, have greatly reduced the differentiation among elite materials. One strategy to overcome this problem is the combination of lines from dissimilar target environments for expansion of genetic diversity among pools (Whitford et al., 2013). In this context, Zhang et al. (2011) investigated the genetic diversity and population structure of 111 cultivars and breeding lines from northern China with 1637 DArT markers to help establish heterotic groups. Such studies provide valuable information for the selection of suitable, genetically divergent crossing parents.

Knowledge about the extent of population structure is also essential for the design of genome-wide association studies (GWAS) and genomic selection (GS). Population structure is one of the main factors affecting linkage disequilibrium (LD) and, in GWAS, can cause the identification of spurious marker–trait associations (Flint-Garcia, 2003; Zhang et al., 2010a). Linkage disequilibrium, the nonrandom association of alleles at different loci (Flint-Garcia, 2003), is a key descriptor of the genetic makeup of plant populations. The extent and pattern of LD can reflect signatures of directional selection for genomic regions harboring genes underlying various traits (Qian et al., 2014). Patterns of LD also enable comparisons of allelic variation at chromosome and subgenome level among populations. Moreover, measures of LD decay are essential for estimation of the required quantity and density of markers for GWAS and GS (Bouchet et al., 2012). Numerous studies have highlighted the importance of analyzing LD and genetic diversity in wheat as a basis for further genome-based research and crop improvement. For example, Zhang et al. (2010a) used 245 genome-wide molecular markers to estimate genetic variation and allelic diversity in 205 hard and soft winter wheat types across different breeding programs in the United States. Benson et al. (2012) investigated population structure, LD, and genetic diversity in 251 winter wheat lines from the eastern United States with enriched Fusarium head blight resistance using DArT and sequence tagged site (STS) markers as a basis for subsequent marker-assisted breeding and association analysis. Cabrera et al. (2014) genotyped two soft winter wheat populations with a 9,000-SNP array and used information on LD and genetic diversity to study population stratification, variation, and genome evolution in American soft winter wheat.

Deep investigation of the hexaploid bread wheat genome is notoriously difficult because of its enormous size (~17×10^9 bp); the close relatedness of the A, B, and D subgenomes; and the high proportion (>80%) of repetitive DNA (Gupta et al., 2008). On the other hand, ultrafast DNA sequencing technologies have accelerated molecular marker discovery and applications even in complex genomes like that of wheat (Edwards et al., 2013), and the introduction of high-throughput genotyping platforms has highly improved this bottleneck and enabled the detailed investigation of genetic material for breeding purposes (Ganal et al., 2012).

In this study, we investigated population genetic parameters on a genome-wide scale among an international diversity set comprising 460 hexaploid wheat accessions using a high-density SNP genotyping array (described by Wang et al., 2014). The population was assembled by commercial wheat breeders with the intention to replenish genetic variation in European breeding pools and to introgress novel resistances into registered elite cultivars. The data were used to assess population substructure of the collection and identify germplasm groups carrying novel diversity for breeding. Furthermore, the use of high-density, genome-wide markers enabled detailed estimation of genetic variation, substructural differentiation, and LD decay patterns within and between subpopulations on a whole-genome, subgenomic, and chromosomal level.

Materials and Methods

Plant Material

We used an international collection of 460 hexaploid wheat accessions representing different geographic origins in China, Europe, North America, and Australia (Supplemental Table S1). Seeds were obtained from W. von Borries-Eckendorf GmbH & Co. KG (Leopoldshöhe, Germany), Wiersum Plant Breeding (Dronten, Netherlands), CSIRO (Canberra, ACT, Australia), and from our own germplasm collection.

The genotype panel includes registered elite varieties, mainly of European and Chinese origin, as well as landraces from China and experimental lines from ongoing breeding programs. In addition to information about pedigree and origin, the genotypes were characterized on the basis of their growth habit into spring wheat forms and early- or late-flowering winter wheat forms. All genotypes were either doubled haploids or selected at the F_6 generation or higher, hence all accessions in the panel were considered to be homozygous.

DNA Extraction and Genotyping

Leaf samples of all genotypes were taken at seedling stage and frozen in liquid nitrogen. Genomic DNA was extracted using a BioSprint 96 magnetic bead robot

extraction system (Qiagen) using the Qiagen DNA Plant Kit and the procedure recommended by the manufacturer. For genome-wide marker analysis, DNA samples of all lines were genotyped using the 90,000-SNP wheat genotyping array (Illumina Inc.) described by Wang et al. (2014), which carries 81,587 functional and valid SNPs. Genotyping was outsourced to TraitGenetics GmbH (Gatersleben, Germany) and automated SNP scoring used a cluster file based on worldwide material described by Wang et al. (2014). Raw marker data was processed by first excluding all markers with more than two called alleles, more than 10% missing data, or minor allele frequency (MAF) less than 10%. This resulted in a total of 22,377 high-quality, polymorphic SNPs in the 450 genotypes that were used for population-structure analyses. For all analyses requiring positional information, we used a set of 18,681 SNPs with MAF ≥5% and known map positions on the consensus map described by Wang et al. (2014).

Population Structure and Genetic Differentiation

Unless otherwise stated, all computations were conducted in the program R (R Development Core Team, 2015). Population structure of the global diversity panel was investigated using different methods. Initially, we applied a Bayesian model-based clustering algorithm, implemented in the program STRUCTURE 2.3.4, to determine the probable number of subpopulations (K). For this we used a representative subset of 2237 markers, comprising every 10th SNP from the 22,377 with MAF ≥10%. A hypothetic K-range was set from two to nine. For each run, burn-in time and replication number were both set to 10,000, with eight runs for each K. The optimum K was then calculated using the ΔK-method of Evanno et al. (2005), implemented in STRUCTURE HARVESTER (Web v0.6.94; Earl and von Holdt, 2012).

Genetic distances were calculated using the modified Roger's distances (MRD) (Wright 1978) based on the 22,377 polymorphic markers with MAF ≥10%. To visualize relationships among the 450 genotypes, principal component analysis (PCA) using the first four principal components and unweighted-pair-group method with arithmetic mean (UPGMA) clustering were performed based on MRD. Subsequently, the genotypes were assigned to groups by applying a k-means clustering approach, using the algorithm of Hartigan and Wong (1979). We first determined an appropriate cluster number by plotting k-means cluster values ranging from 1 to 15 against their corresponding within-cluster sum of squares (Hartigan and Wong, 1979). The k-means cluster value at which a clear bend appears in the curve is considered as an appropriate cluster number. The k-means clustering was conducted with a start value for random sets of 10.

A heat map of genetic relatedness was drawn by plotting UPGMA trees with related k-means cluster assignments against each other. To further investigate population structure within subgroups, unrooted phylogenetic trees were plotted for each detected subpopulation. The degree of genetic relatedness among the tested accessions was calculated as the average identity-by-state (IBS) using 18,681 mapped, polymorphic markers (MAF ≥5%) for the whole population and the subpopulations separately, while high IBS values indicate strong kinship (Aulchenko et al., 2007).

To analyze and compare the level of genetic variability among the subgroups on a whole-genome and subgenome level, the gene diversity over all loci, also designated as expected heterozygosity, was calculated using the method invented by Nei (1973). The gene diversity ranges from zero to one and peaks when many alleles are at equal frequencies. Additionally, to measure the degree to which genetic differentiation in the whole panel can be explained by differentiation within subpopulations, Wright's fixation index (F_{ST}) was computed using the GENEPOP 4.2.2 software (Raymond and Rousset, 1995; Rousset, 2008).

Identification and Characterization of Artificial Selection Signatures

To identify loci under directional selection between or within the different subpopulations, we used the F_{ST} outlier detection method implemented in the LOSITAN workbench (Antao et al., 2008). The calculation was performed using an infinite allele model with 100,000 simulations. All loci that fell outside of the 95% confidence interval were assumed to be candidates for directional selection and used for further analysis. Because only a few published studies report marker–trait associations using the recently generated wheat 90,000-SNP genotyping array and no full reference genome aligning these SNP markers to gene annotations is available, we compared the target regions to information from literature to investigate the potential underlying functions of strongly selected loci. Particular attention was paid to marker–trait associations, biparental quantitative trait loci (QTL), and meta-QTL described for grain yield (Zhang et al., 2010b), plant height, baking quality (http://ccg.murdoch.edu.au), disease resistance (Klahr et al., 2007; Löffler et al., 2009), and flowering-time or vernalisation-related traits (Zanke et al., 2014).

Analysis of Linkage Disequilibrium

As a measure of LD between SNP markers we calculated r^2 (Hao et al., 2007) between intrachromosomal marker pairs. To describe the relationship between LD decay and genetic map distance, a locally estimated scatterplot smoothing (LOESS) curve with a smoothing degree of 0.3 was calculated. The threshold for LD was comparatively set to $r^2 = 0.1$ and $r^2 = 0.2$, assuming that r^2 between marker pairs above these values is likely to be caused by genetic linkage. The intersection of the LOESS curve with the threshold line was assumed to be the estimate for the degree of LD. Comparative analysis was performed separately within all subpopulations for markers mapped on chromosomes representing the A, B, and D subgenomes, respectively. Marker pairs with a distance above 50 cM were assumed to be unlinked (Nielsen et al., 2014) and therefore not considered in the LD decay estimation.

As an estimate of the selection pressure on potential candidate loci for directional selection, mean r^2 values were calculated in 2-cM windows around SNPs showing significant selection. For regions harboring more than one marker, the mean LD was calculated for the whole section plus one additional cM on each flanking side. For detailed investigation of the extent and pattern of LD in particular genomic regions, LD heat maps were plotted for adjacent marker pairs.

Results

Single Nucleotide Polymorphism Genotyping and Marker Distribution

Single nucleotide polymorphism genotyping of 460 wheat lines with the 90,000-SNP Infinium array provided genotype calls for a total of 81,587 SNPs. Of these, we excluded 17.9% SNPs with more than 10% missing values and 66.6% that were either monomorphic or showed a MAF <10%. Ten genotypes with more than 10% missing values were also excluded, resulting in 22,377 high-quality SNPs with MAF ≥10% and 24,684 with MAF ≥5% across 450 wheat accessions. These markers were used in the subsequent population structure analyses.

Of these SNPs, 18,681 with MAF ≥5% (16,206 of which have MAF ≥10%) have known unique positions on the consensus map of Wang et al. (2014). The markers cover all 21 *T. aestivum* chromosomes, with a total of 7312, 9495, and 1874 markers on the A, B, and D subgenome chromosomes and an average distribution of one marker per 0.16, 0.12, and 1.15 cM, respectively. As summarized in Supplementary Table S2, the D subgenome shows the highest number of large marker gaps, with two regions on chromosomes 4D and 7D being larger than 30 cM. Additionally, five segments with no mapped markers were identified on chromosomes 1D, 4D, 5D, and 6D ranging in size from 20 to 30 cM, whereas only one gap of this size could be identified on the A subgenome (on chromosome 7A). The largest chromosome region with no polymorphic SNPs in the B subgenome was a 15.77-cM segment on chromosome 5B. In total, 20, 23, and 68 sections between 5 and 20 cM without polymorphic SNPs were found in the A, B, and D subgenomes, respectively.

Population Structure and Genetic Relatedness

By applying the rate of change in the Napierian logarithm probability relative to standard deviation (ΔK) (Evanno et al., 2005), the Bayesian clustering model implemented in the STRUCTURE 2.3.4 software found an optimum of $K = 3$ subpopulations. In contrast, the within-sum of squares curve for a hypothetical cluster number ranging from 1 to 15 showed a clear bench at $k = 5$, suggesting that there are five main subgroups within the whole population. Correspondingly, subsequent PCA-based *k*-means clustering further separated the two bigger clusters found by STRUCTURE into two subgroups, respectively (Fig. 1). Furthermore, the heat map of relatedness revealed a distinct structure in the two larger groups into at least two major clusters, whereas no strong genetic relations could be found in the third cluster (Fig. 2). Based on these consistent findings, the five subpopulations (SPs) SP1 ($n = 32$), SP2 ($n = 85$), SP3 ($n = 115$), SP4 ($n = 44$), and SP5 ($n = 174$) were used for further analyses.

The subgroup assignments for all genotypes within the diversity panel are summarized in Table 1. These generally match available information on origin and growing habit. Subpopulations 1 and 2 solely consist of Chinese material, while SP4 and SP5 include most of the European lines, with the exemption of seven genotypes in SP5 that were assumed to have a Chinese genetic background (4, 5, 6, 10, 12, 17, and 18). The vast majority of European commercial elite varieties are grouped in SP5, while SP4 contains mostly breeding lines with a strong European genetic background, including the German elite cultivars SU Anapolis and Forum. Subpopulation 3 comprises wheat accessions from different origins, such as breeding lines from crosses between European varieties with lines from CIMMYT, Chinese, American, and Australian material and three registered cultivars from Europe (Intro, Linus, and Florence Aurore). Subpopulation 3 is therefore considered to be a mixed subgroup. The distribution of the diverse genotypes based on the population structure analysis methods furthermore resulted in a clear separation of late-flowering winter wheat lines from spring wheat and early-flowering winter wheat accessions. Subpopulations 1, 2, and 3 consisted almost exclusively of spring wheat and early-flowering winter wheat types. Only 21 genotypes described as late-flowering winter wheat fell into these subgroups. On the other hand, SP4 and SP5 included predominantly late-flowering winter wheat genotypes with the exemption of 11 and 37 early-flowering lines, respectively (Supplemental Table S1).

To further investigate the population structure and genetic relationships within the subgroups, dendrograms were plotted (Supplemental Fig. S1) for all five subpopulations. The unrooted phylogenetic trees reveal a certain amount of structure within the five subpopulations and give more detailed insight into the degree of relatedness among the genotypes in each cluster. The gene diversity of the whole population and the different subgroups on the chromosome and subgenome level is summarized in Supplemental Table S3, while pairwise F_{ST} values between the five subpopulations are shown in Table 2. Supplemental Fig. S2 compares minor allele frequencies of polymorphic SNP markers in the whole population to the five subpopulations. The IBS values, a measure for genetic kinship of the genotypes, are displayed as boxplots for the different subgroups in Fig. 3.

Linkage Disequilibrium Decay

The results of intrachromosomal pairwise LD analysis in the A, B, and D subgenomes, respectively, compared with subgenome LD in the whole population, are shown in

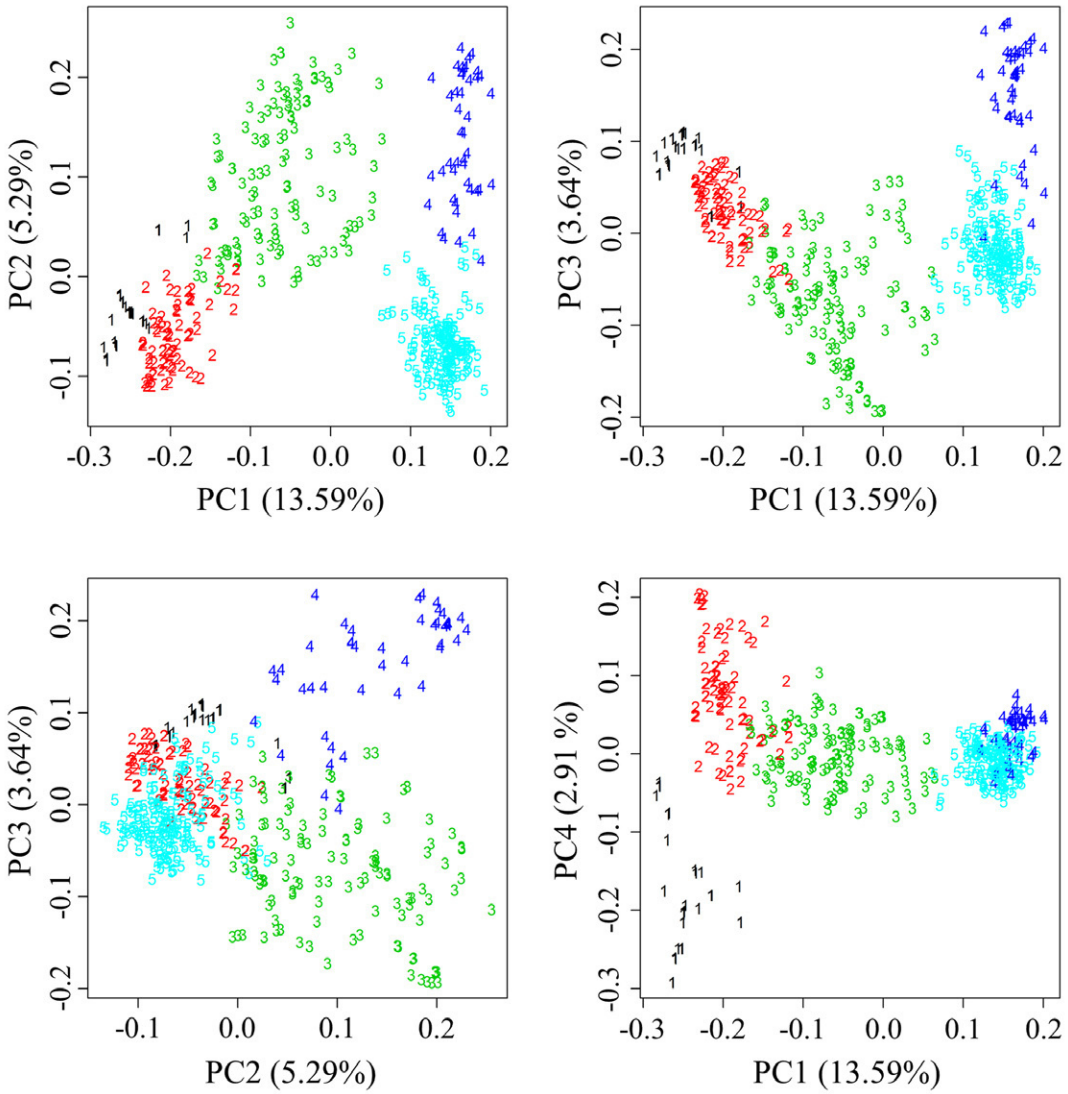


Figure 1. Principal component analysis (PCA)-based *k*-means clustering. Calculations are based on modified Roger's distances for 22,377 polymorphic single nucleotide polymorphism (SNP) markers (minor allele frequency ≥10%) and 450 genotypes. Different numbers and colors indicate the five different clusters found for the four principal components (PC1–PC4).

Supplemental Table S3 and Fig. 4. The whole-genome average LD varies strongly between the whole set and the different subpopulations, with r^2 being highest in SP1 (0.296) and the lowest in SP5 (0.064). Overall, the D subgenome showed the highest r^2, followed by the B and D subgenomes, respectively. An exception to this general observation was seen for SP2, which had higher average LD in the A subgenome (0.113) than in the B subgenome (0.093).

As shown in Fig. 4, LD in the whole population decayed the fastest in the A subgenome, reaching r^2 values for intrachromosomal marker pairs below 0.1 at 4.1 cM. The B genome LD decayed 0.5 cM subsequently, whereas r^2 measures for the D genome did not fall under 0.1 until 11.2 cM.

Comparing the LD decay of the subpopulations in the three subgenomes revealed that the highest extent of LD was present in SP1 over all three subgenomes, with average LD decay distances of 54.7 and 40.8 cM (r^2 = 0.1) for the A and D subgenomes, respectively. In the B subgenome, r^2 values never reached the bottom line of 0.1. Subpopulation 4 showed a comparatively high extent of LD in the B subgenome, with a decay distance of 14.5 cM, while LD decayed considerably faster in SP2, SP3, and SP5 (5.8, 5.3, and 7 cM, respectively). The LD in the D subgenome decayed next fastest in SP4 (11.7 cM), followed by SP3 (12.7 cM), SP2 (14 cM), and SP5 (14.7 cM).

Quantitative Trait Loci and Metaquantitative Trait Loci for Key Breeding Targets Define Selection Signatures in Distinct Wheat Genepools

A total of 150 loci located on 12 chromosomes of the A, B, and D subgenomes (four on each subgenome) were found to exhibit signatures of directional selection in one or more of the five subpopulations defined by the structure analysis as identified by F_{ST} outlier detection (Supplemental Table S5). As shown in Fig. 5, many of these selective sweeps have led to absolute or near fixation for specific

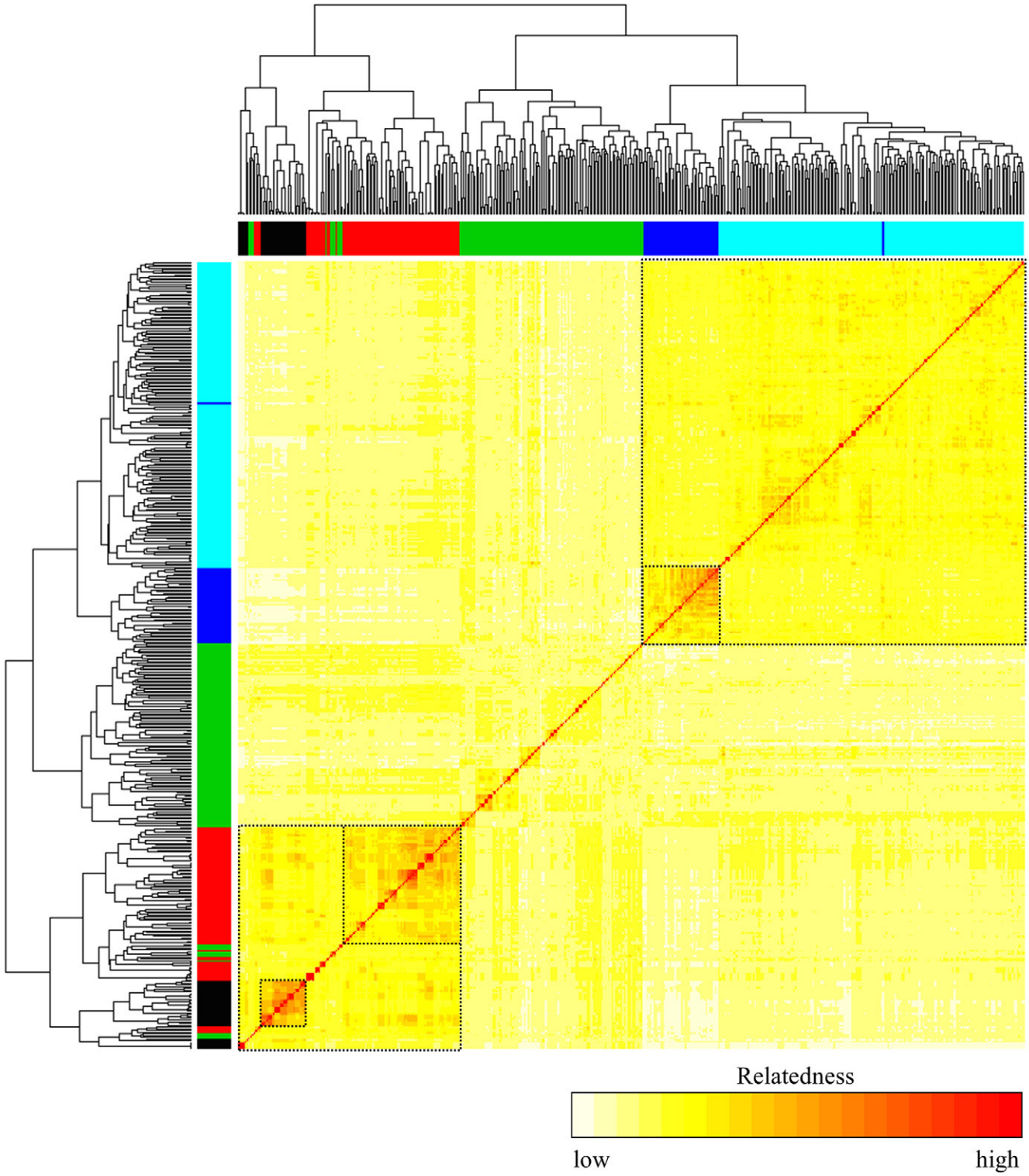


Figure 2. Heat map of relatedness among 450 genotypes. Dendrograms were plotted using unweighted-pair-group method with arithmetic mean (UPGMA) clustering based on modified Roger's distances for 22,377 polymorphic single-nucleotide-polymorphism (SNP) markers (minor allele frequency ≥10%). Degree of relatedness is indicated by colors from light yellow (no relatedness) to red (strong relatedness). Colors next to UPGMA trees correspond to the *k*-means clustering assignment.

alleles at the respective loci in one or more of the subpopulations. Other loci show less pronounced directional selection but with similar patterns in multiple subpopulations. Supplementary Table S4 summarizes the mean r^2 values within candidate regions for directional selection.

Table 1. Pairwise Wright's fixation index (F_{ST}) between the five subpopulations (SP) in a panel of 460 genetically diverse wheat accessions.

F_{ST} values	SP1	SP2	SP3	SP4
SP2	0.2586			
SP3	0.2504	0.1591		
SP4	0.5537	0.4486	0.2443	
SP5	0.4373	0.3244	0.1798	0.2232

Table 2. Subpopulation (SP) characteristics for a panel of 460 genetically diverse wheat accessions.

	SP1	SP2	SP3	SP4	SP5
Origin	Chinese	Chinese	78% Chinese, 22% European–North American–Australian	European	European
Growth habit	Spring wheat	Spring wheat and early-flowering winter wheat	Spring wheat and early-flowering winter wheat,	Late-flowering winter wheat	Late-flowering winter wheat

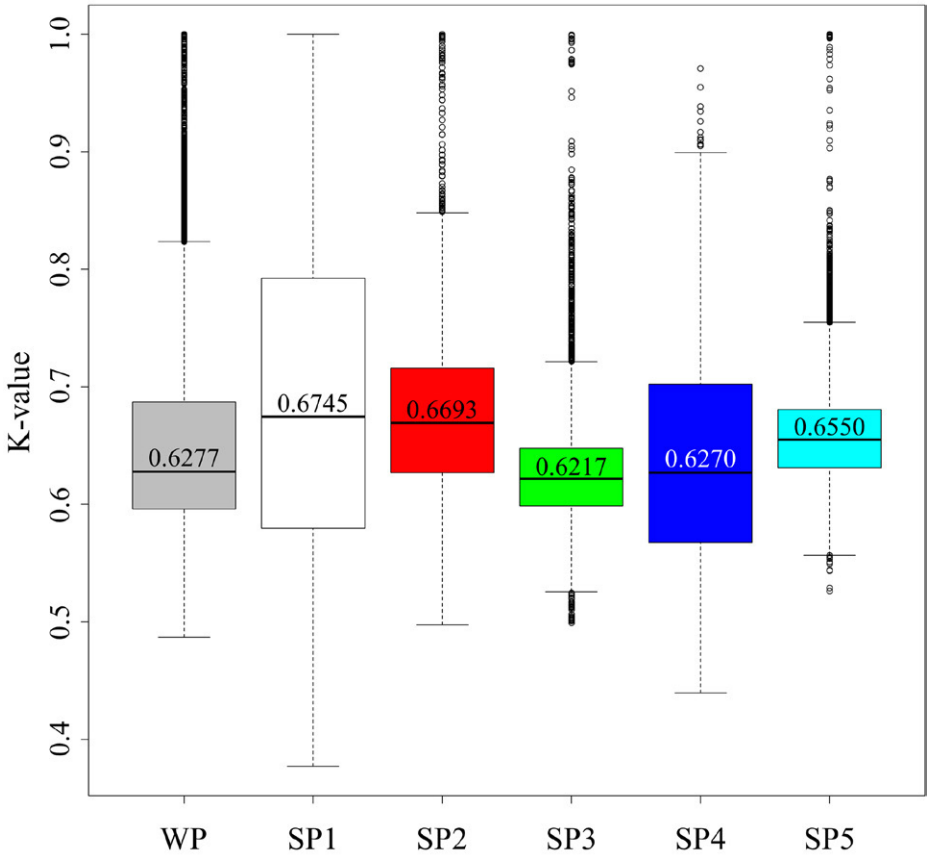


Figure 3. Average pairwise identity-by-state (IBS) estimates for the whole population (WP) and subpopulations (SP). High IBS values indicate high kinship among the wheat lines in the corresponding group.

Heat maps drawn to visualize and characterize LD in selected chromosome regions with particularly prominent F_{ST} outliers (Fig. 6) revealed that the majority of SNPs located in these chromosomal regions are involved in distinct LD blocks, underlining the assumption that these loci have undergone strong directional selection. Surprisingly, 42% of the candidate loci were located in a single 10-cM segment on chromosome 1B, hence, we investigated this region in more detail to determine its background in terms of known QTL for important agronomic traits (Supplemental Table S6) and its diversity across the different wheat subpopulations. Closer investigation revealed an enormous LD block, particularly in subpopulation SP4 (Fig. 6), whose members share recent ancestry from crosses with the accession 'Shanghai-3/Catbird' (SHA3/CBRD) developed by CIMMYT. The accession SHA3/CBRD carries a 1B/1R translocation from rye that confers multiple positive traits but is known to suppress recombination on chromosome 1B.

Discussion

The exploitation of genetic resources remains the most promising option to accelerate wheat improvement and to enhance genetic variation for selection (Mohammadi and Prasanna, 2003). Indeed, enrichment of modern gene pools depleted by intensive breeding (Fu et al., 2006) is vital to reverse the erosion of genetic variability and further improve yield and heterosis potential. A detailed genetic description of potential crossing partners in breeding programs, using the latest low-cost, high-density genotyping tools, opens the possibility to

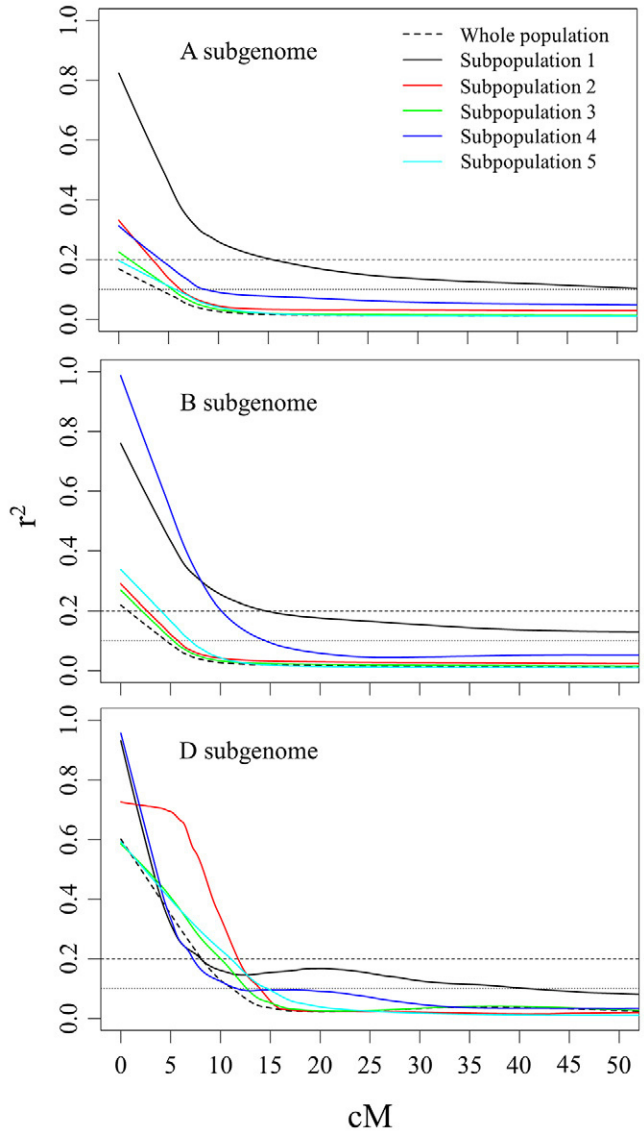


Figure 4. Linkage disequilibrium (LD) decay comparison between the different subgroups in the A, B, and D subgenomes, respectively. The LD decay is indicated as the intersection of the locally estimated scatterplot smoothing (LOESS) curves with the dashed lines (standard cutoff, r^2 = 0.1/0.2). The r^2 values were calculated with polymorphic single nucleotide polymorphism markers (minor allele frequency ≥5%).

precisely reinstate depleted diversity in genomic target regions or at a subgenome or whole-genome level. In this study, we describe the population structure and genome-wide diversity in an international collection of 460 hexaploid bread wheat accessions using the recently released 90,000-SNP Illumina Infinium wheat genotyping array. By implementing genetic linkage data from almost 20,000 polymorphic SNPs that have known consensus map positions, we were able to perform a detailed analysis of local, subgenomic, and genome-wide LD. The resulting dataset provides unique insight into the genomic consequences of artificial selection in a comprehensive international collection of bread wheat. The results provide important information for breeders attempting to improve recombination of strongly conserved LD blocks in low-diversity genome regions associated with signatures of selection for key adaptive and agronomic traits.

High-Density Single Nucleotide Polymorphism Genotyping and Marker Distribution

The 90,000 Illumina Infinium SNP Array is a recent, high-density wheat genotyping platform that was first described by Wang et al. in 2014. To our knowledge, our study is the first to implement the 90,000-SNP array for comprehensive analysis of genome-scale diversity in a large international panel of unrelated wheat genotypes. Our results therefore provide valuable information on performance and applicability of the array for the design of GWAS or GS approaches in bread wheat.

A number of previous studies have shown the value of genotyping-by-sequencing (GBS) approaches for high-density SNP scoring in wheat (Poland et al., 2012a,b). Although GBS has obvious merits in terms of potentially low cost and discovery of unknown alleles (Elshire et al., 2011), the complex data analysis pipeline required to successfully apply GBS in plants with polyploid genomes like wheat, particularly in nonrelated populations where segregation data cannot assist in locus calling, demand considerable bioinformatics expertise and infrastructure. Furthermore, the intellectual property associated with GBS technologies complicate its use by commercial breeders, who often prefer standardized array-based data formats for marker-assisted selection, GWAS, and GS.

The low proportion of polymorphic SNPs we detected on D subgenome chromosomes is consistent with previous findings (Cabrera et al., 2014; Zanke et al., 2014; Zegeye et al., 2014) and reflects how the evolutionary bottleneck in the D subgenome of bread wheat (Wang et al., 2014) has caused ascertainment bias for polymorphic SNPs from the different subgenomes (Würschum et al., 2013; Thomson, 2014). Compared with previous studies that used a 9,000-SNP array or GBS to analyze hexaploid wheat, the 90,000-array, to some extent, alleviates the difficulty of D subgenome genotyping. In total, between 52 and 607 polymorphic SNPs were located on the seven D-subgenome chromosomes in our study. This elevated level of diversity, compared with many previous studies, reflects both the larger size and the more diverse nature of our population. Nevertheless, the very low overall genetic variation in the D subgenome of hexaploid wheat, even among very distant gene pools, underlines the pressing need for introduction of new D-subgenome diversity via synthetic wheat lines (Henry and Nevo, 2014; Jia et al., 2013).

Chromosome-Scale and Subgenome Diversity Patterns

The low-diversity D subgenome includes the highest number of marker gaps greater than 5 cM in which no polymorphic SNPs could be mapped (Supplemental Table S2), along with the two largest nonpolymorphic chromosomal sections (>30 cM), on 4D and 7D. Since

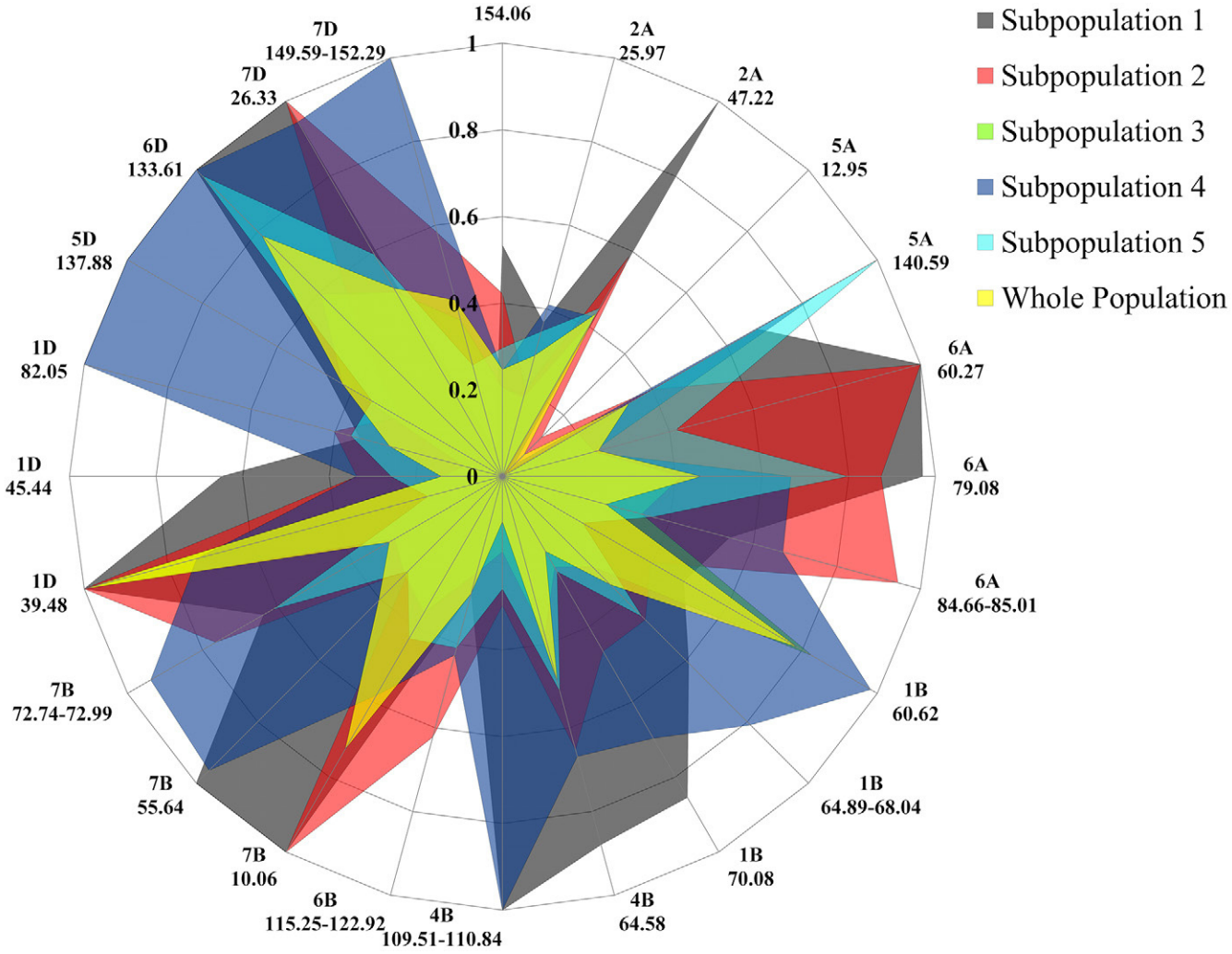


Figure 5. Average linkage disequilibrium (LD) at 24 Wright's fixation index (F_{ST}) outlier loci compared over the whole population and the five subgroups, respectively. Candidate loci were calculated using the LOSITAN workbench (Antao et al., 2008) and represent potential targets for directional selection in the corresponding germplasm groups. The LD was calculated as r^2 for single nucleotide polymorphism markers with a minor allele frequency ≥5%.

the SNPs on this genotyping array were mainly developed from protein-coding transcriptome sequences (Wang et al., 2014), the enormous number of repetitive, noncoding DNA sequences (>80%) throughout the wheat genome (Gupta et al., 2008) certainly contribute to variation in chromosomal distribution of these SNPs. On the other hand, SNPs targeting coding regions of the genome are particularly useful for GWAS and GS approaches, hence, information about the distribution of these SNPs are highly relevant for the design of GWAS experiments and high-resolution QTL mapping and cloning.

Population Structure and Genetic Relatedness

Clear population stratification was identified between subpopulations comprising Chinese (SP1 and SP2) and European materials (SP4 and SP5). Balfourier et al. (2007) report a similar clear separation between Asian and European wheat lines based on SSR marker data. In our study, seven lines originating from China that clustered closely together with European registered cultivars in SP5 are likely to represent interpool germplasm exchanges by Chinese breeders. Subpopulations 1, 2, and 3 consist almost exclusively of spring and early-flowering winter wheat types, whereas SP4 and SP5 comprise mainly late-flowering winter wheat accessions. The clear ecogeographic subdivision between spring and winter wheat has been well documented previously (e.g., Chao et al., 2010). Within these material groups, however, the phylogenetic trees and heat maps of relatedness for the five subpopulations provide highly beneficial information for choosing genetically divergent crossing partners within geographically limited germplasm.

Gene Diversity and Genetic Differentiation

Analysis of gene diversity over all loci revealed strong differences in genetic variability between and within the subpopulations on the whole-genome, subgenome, and single-chromosome level. Considerable variation was observed within the different subgroups for diversity of chromosomes from the same subgenome. As expected

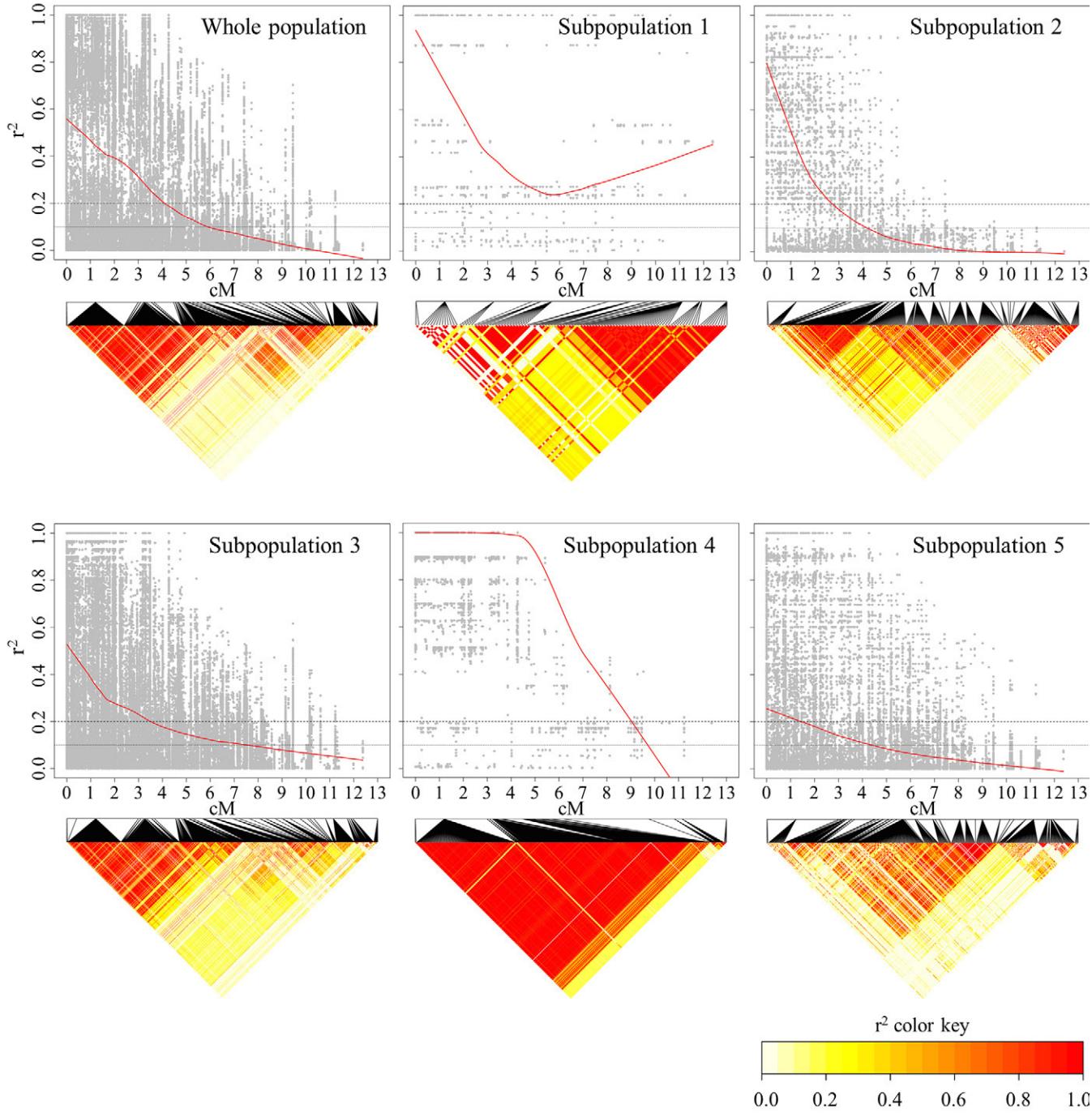


Figure 6. Comparison of the decay of linkage disequilibrium (LD) on chromosome 1B between 59 and 72 cM in the whole population and the five subpopulations. The plots show the degree of LD decay, while the colored heat maps represent the corresponding pairwise LD patterns in this chromosome region. The LD was calculated as r^2 for polymorphic single nucleotide polymorphism markers with a minor allele frequency ≥5%.

from previous studies, the D genome showed the lowest genetic diversity for all five subpopulations (Akhunov et al., 2009, 2010; Hao et al., 2011; Cabrera et al., 2014), reflecting the most recent polyploidy bottleneck of hexaploid wheat (Marcussen et al., 2014; Chao et al., 2010). On the other hand, in some subpopulations individual chromosomes showed considerably higher gene diversity than the D subgenome average of 0.32, particularly chromosome 3D in SP4, which showed gene diversity of 0.47. In contrast, Akhunov et al. (2010) found a distinctly lower diversity on chromosome 3D than on the other D-subgenome chromosomes. Wang et al. (2014) also report only moderate to low diversity on chromosome 3D, especially compared with the most diverse D-subgenome chromosomes 1D and 2D. These contrasting results suggest that the high diversity on 3D in SP4 in our study should be further investigated as a potentially promising resource for genetic diversity on this chromosome.

The comparatively high A- and B-subgenome diversity in SP5, which contains most registered elite varieties within our diversity set, is considerably higher than reported by Nielsen et al. (2014) for 94 European elite wheat varieties. This might be explained by a higher polymorphic information content (PIC) from the genome-wide SNPs compared with the DArT markers used in the previous study. On the other hand, Mackay et al. (2014) also report a lower average PIC of 0.251 for 64 registered varieties from the United Kingdom genotyped with the same 90,000-SNP array, whereas Würschum et al. (2013) reported a PIC of 0.33 for a set of 172 European cultivars analyzed with a 9,000-SNP chip. The higher gene diversity in SP5 in our study possibly reflects recent breeding progress in diversification of this gene pool. In contrast to SP5, the smaller European subgroup SP4 had a distinctly lower average gene diversity (0.34), comparable to those in the studies of Mackay et al. (2014) and Würschum et al. (2013). However, the highly related pedigree of SP4, whose members derived mainly from a single CIMMYT donor, certainly contributes to its the low genetic variation. Gene diversity in the Chinese subgroup SP2 (0.39) reflects the level of 0.40 found in similar materials by Zhang et al. (2011), whereas the much smaller group SP1 was considerably less diverse (0.28). This may be because smaller populations tend to expose less gene diversity than larger ones (Würschum et al., 2013). The marker data revealed that the Chinese landrace Wangshubai was represented three times in SP1, possibly due to declaration errors, meaning that SP1 contained only 30 distinct genotypes.

Subgenomic Signatures of Directional Selection

Knowledge about the extent and pattern of LD in plant populations is a crucial prerequisite for genome-based plant research and can support breeding in various ways. The two major uses of LD in plants are (i) to analyze marker–trait associations as LD determines the resolution of a GWAS experiment (Flint-Garcia, 2003; Benson et al., 2012) and (ii) to investigate genetic diversity in natural populations or germplasm collections for the study of population genetics in crop breeding programs (Gupta, 2008). Strong mean LD in the D genome has previously been reported by several studies (Chao et al., 2010; Hao et al., 2011; Chen et al., 2012; Nielsen et al., 2014; Wang et al., 2014); however, investigations that pinpoint specific wheat chromosomes and chromosome regions with the strongest signatures of selection are less common. We demonstrated that all bread wheat subgenomes carry chromosome regions that have been subject to strong directional selection in specific subpopulations with distinct ecogeographic distributions. On the other hand, our data revealed that the extent and decay of LD is highly dependent on population structure within the different subgenomes.

Knowledge about local LD in different breeding pools is an important prerequisite for rejuvenation of diversity within subgenomic signatures of selection at essential adaptation, plant height, or grain quality loci. Background selection with high-density markers, using donors selected based on the data presented here, might help increase levels of recombination around causative genes for significant selective sweeps.

Conclusions for Applied Breeding

Exploitation of genetic resources is vital to broaden genetic variation for wheat improvement and to break present yield barriers. Our study discloses genetic relationships and diversity in a broad international panel of hexaploid wheat lines and suggests that there is a suitable variation among pools, especially between spatially distinct groups, for targeted introgression of useful variation with suitable recombination into diversity-poor regions of elite varieties. High-density SNP genotype data for large diversity collections can support the selection of crossing parents both between and within gene pools to increase overall genetic variation in a breeding program. Our study provides valuable basic information for combining materials with similar adaptations that show a high phenotypic and genotypic variation to broaden the diversity for selection.

Chinese wheat materials represent a particularly interesting resource for transfer of quantitative disease resistances into depleted European gene pools. For example, the diversity set used in the present study includes some of the most important donors (e. g. cultivars Sumai 3, Wangshubai, Ning 7840, Dream, and Lynx) of resistance to Fusarium head blight, one of the most important diseases of bread wheat worldwide (Burstmayer et al., 2009). Use of the genotype data to perform GWAS for such resistances can provide important knowledge about the genetic basis of important disease resistances and facilitate the deployment of introgressed resistances in cultivars that fulfill the requirements of quality and local adaptation.

Acknowledgments

This study was performed with funding from the German Federal Ministry of Food and Agriculture under coordination of the Federal Agency of Renewable Resources (FNR), Gülzow, Germany. We also thank Stavros Tzigos for excellent technical assistance and Bertrand Schuiling, Wiersum Plant Breeding, Netherlands, for providing part of the diversity collection.

References

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

Akhunov, E., C. Nicolet, and J. Dvorak. 2009. Single nucleotide polymorphism genotyping in polyploidy wheat with the Illumina GoldenGate assay. Theor. Appl. Genet. 119:507–517. doi:10.1007/s00122-009-1059-5

Antao, T., A. Lopes, R.J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}–outlier method. BMC Bioinformatics 9:323. doi:10.1186/1471-2105-9-323

Aulchenko, Y.S., S. Ripke, A. Isaacs, and C.M. Duijn. 2007. GenABEL: An R library for genome-wide association analysis. Bioinformatics 23:1294–1296. doi:10.1093/bioinformatics/btm108

Balfourier, F., V. Roussel, P. Strelchenko, F. Exbrayat-Vinson, P. Sourdille, G. Boutet, J. Koenig, C. Ravel, O. Mitrofanova, M. Beckert, and G. Charmet. 2007. A worldwide bread wheat core collection arrayed in a 384-well plate. Theor. Appl. Genet. 114:1265–1275. doi:10.1007/s00122-007-0517-1

Benson, J., G. Brown-Guedira, J.P. Murphy, and C. Sneller. 2012. Population structure, linkage disequilibrium, and genetic diversity in soft winter wheat enriched for Fusarium Head Blight resistance. Plant Gen. 5:71–80. doi:10.3835/plantgenome2011.11.0027

Bouchet, S., D. Pot, M. Deu, J.F. Rami, C. Billot, X. Perrier, R. Rivallan, L. Gardes, P. Wenzl, A. Kilian, and J.C. Glaszmann. 2012. Genetic structure, linkage disequilibrium and signature of selection in sorghum: Lessons from physically anchored DArT markers. PLoS ONE 7:E33470. doi:10.1371/journal.pone.0033470

Burstmayer, H., T. Ban, and J.A. Anderson. 2009. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. Plant Breed. 128:1–26. doi:10.1111/j.1439-0523.2008.01550.x

Brisson, N., P. Gate, D. Gouache, G. Charmet, F.X. Oury, and F. Huard. 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crops Res. 119:201–212. doi:10.1016/j.fcr.2010.07.012

Cabrera, A., E. Souza, M. Guttieri, A. Sturbaum, A. Hoffstetter, and C. Sneller. 2014. Genetic diversity, linkage disequilibrium, and genome evolution in soft winter wheat. Crop Sci. 54:2433–2448. doi:10.2135/cropsci2013.09.0601

Chao, S., J. Dubcovsky, J. Dvorak, M.C. Luo, S.P. Baenziger, R. Matnyazov, D.R. Clark, L.E. Talbert, J.A. Anderson, S. Dreisigacker, K. Glover, J. Chen, K. Campbell, P.L. Bruckner, J.C. Rudd, S. Haley, B.F. Carver, S. Perry, M.E. Sorrells, and E.D. Akhunov. 2010. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11:727. doi:10.1186/1471-2164-11-727

Chen, X., D. Min, T.A. Yasir, and Y.G. Hu. 2012. Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. PLoS One 7: e44510. doi:10.1371/journal.pone.0044510

Earl, D.A., and B.M. von Holdt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Gen. Resour. 4:359–361. doi:10.1007/s12686-011-9548-7

Edwards, D., J. Batley, and R.J. Snowdon. 2013. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. 126:1–11. doi:10.1007/s00122-012-1964-x

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. doi:10.1371/journal.pone.0019379

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Flint-Garcia, S.A. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357–374. doi:10.1146/annurev.arplant.54.031902.134907

Fu, Y.B., G.W. Peterson, J.K. Yu, L. Gao, J. Jia, and K.W. Richards. 2006. Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. Theor. Appl. Genet. 112:1239–1247. doi:10.1007/s00122-006-0225-2

Ganal, M.W., A. Polley, E.M. Graner, J. Plieske, R. Wieseke, H. Luerssen, and G. Durstewitz. 2012. Large SNP arrays for genotyping in crop plants. J. Biosci. 37:821–828. doi:10.1007/s12038-012-9225-3

Gupta, P.K., R.R. Mir, A. Mohan, and J. Kumar. 2008. Wheat genomics: Present status and future prospects. Int. J. Plant Genomics 2008:896451. doi:10.1155/2008/896451

Hao, C., L. Wang, H. Ge, Y. Dong, and X. Zhang. 2011. Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) revealed by SSR markers. PLoS ONE 6:e17279. doi:10.1371/journal.pone.0017279

Hao, K., X. Di, and S. Cawley. 2007. LdCompare: Rapid computation of single- and multiple-marker r^2 and genetic coverage. Bioinformatics 23:252–254. doi:10.1093/bioinformatics/btl574

Hartigan, J.A., and M.A. Wong. 1979. A k-means clustering algorithm. Series C. Appl. Stat. 28:100–108. doi:10.2307/2346830

Henry, R.J., and E. Nevo. 2014. Exploring natural selection to guide breeding for agriculture. Plant Biotechnol. J. 12:655–662. doi:10.1111/pbi.12215

Jia, J., S. Zhao, X. Kong, Y. Li, G. Zhao, W. He, R. Appels, M. Pfeifer, Y. Tao, X. Zhang, Ruilian J., C. Zhang, R. Jing, C. Zhang, Y. Ma, L. Gao, C. Gao, M. Spannagl, K.F.X. Mayer, D. Li, S. Pan, F. Zheng, Q. Hu, X. Xia, J. Li, Q. Liang, J. Chen, T. Wicker, C. Gou, H. Kuang, G. He, Y. Luo, B. Keller, Q. Xia, P. Lu, J. Wang, H. Zou, R. Zhang, J. Xu, J. Gao, C. Middleton, Z. Quan, G. Liu, J. Wang, International Wheat Genome Sequencing Consortium, H. Yang, X. Liu, Z. He, L. Mao, and J. Wang. 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95. doi:10.1038/nature12028

Klahr, A., G. Zimmermann, G. Wenzel, and V. Mohler. 2007. Effects of environment, disease progress, plant height and heading date on the detection of QTLs for resistance to Fusarium head blight in an European winter wheat cross. Euphytica 154:17–28. doi:10.1007/s10681-006-9264-7

Longin, C.F.H., and J.C. Reif. 2014. Redesigning the exploitation of wheat genetic resources. Trends Plant Sci. 19:631–636. doi:10.1016/j.tplants.2014.06.012

Löffler, M., C.C. Schön, and T. Miedaner. 2009. Revealing the genetic architecture of FHB resistance in hexaploid wheat (*T. aestivum* L.) by QTL meta-analysis. Mol. Breed. 23:473–488. doi:10.1007/s11032-008-9250-y

Mackay, I.J., P. Bansept-Basler, T. Barber, A.R. Bentley, J. Cockram, N. Gosman, A.J. Greenland, R. Horsnell, R. Howells, D.M. O'Sullivan, G.A. Rose, and P.J. Howell. 2014. An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 (Bethesda) 4:1603–1610. doi:10.1534/g3.114.012963

Marcussen, T., S.R. Sandve, L. Heier, M. Spannagl, M. Pfeifer, The International Wheat Genome Sequencing Consortium, K.S. Jakobsen, B.B.H. Wulff, B. Steuernagel, K.F.X. Mayer, and O.D. Olsen. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science 345. doi:10.1126/science.1250092

Melchinger, A.E. 1999. Genetic diversity and heterosis. In: J.T. Gerdes, editor, Genetics and exploitation of heterosis in crops. ASA, CSSA, SSSA, Madison, WI. p. 99–118. doi:10.2134/1999.geneticsandexploitation.c10

Mohammadi, S.A., and B.M. Prasanna. 2003. Analysis of genetic diversity in crop plants: Salient statistical tools and considerations. Crop Sci. 43:1235–1248. doi:10.2135/cropsci2003.1235

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70:3321–3323. doi:10.1073/pnas.70.12.3321

Nielsen, N.H., G. Backes, J. Stougaard, S.U. Andersen, and A. Jahoor. 2014. Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum* L.) varieties. PLoS ONE 9:e94000. doi:10.1371/journal.pone.0094000

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253. doi:10.1371/journal.pone.0032253

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, Y. Dreisigacker, J. Crossa, H. Sanchez-Villeda, M. Sorrells, and J.L. Jannink. 2012b. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen. 5:103–113. doi:10.3835/plantgenome2012.06.0006

Qian, L., W. Qian, and R.J. Snowdon. 2014. Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. BMC Genomics 15:1170 10.1186/1471-2164-15-1170.

Ray, D.K., N. Ramankutty, N.D. Mueller, P.C. West, and J.A. Foley. 2012. Recent patterns of crop yield growth and stagnation. Nat. Commun. 3:1293. doi:10.1038/ncomms2296

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. 86:248–249.

R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Reif, J.C., P. Zhang, S. Dreisigacker, M.L. Warburton, M. van Ginkel, D. Hoisington, M. Bohn, and A.E. Melchinger. 2005. Wheat genetic diversity trends during domestication and breeding. Theor. Appl. Genet. 110:859–864. doi:10.1007/s00122-004-1881-8

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8: 103–106. doi:10.1111/j.1471-8286.2007.01931.x

Thomson, M.J. 2014. High-throughput SNP genotyping to accelerate crop improvement. Plant Breed. Biotechnol. 2:195–212. doi:10.9787/PBB.2014.2.3.195

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

Whitford, R., D. Fleury, J.C. Reif, M. Garcia, T. Okada, V. Korzun, and P. Langridge. 2013. Hybrid breeding in wheat: Technologies to improve hybrid wheat seed production. J. Exp. Bot. 64:5411–5428. doi:10.1093/jxb/ert333

Wright, S. 1978. Evolution and the genetics of populations, Vol. IV. Variability within and among natural populations. University of Chicago Press, Chicago, IL.

Würschum, T., S.M. Langer, C.F.H. Longin, V. Korzun, E. Akhunoc, E. Ebmeyer, R. Sachschneider, J. Schacht, E. Kazmann, and J.C. Reif. 2013. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor. Appl. Genet. 126:1477–1486. doi:10.1007/s00122-013-2065-1

Zanke, C., J. Ling, J. Plieske, S. Kollers, E. Ebmeyer, V. Korzun, O. Argillier, G. Stiewe, M. Hinze, S. Beier, M.W. Ganal, and M.S. Röder. 2014. Genetic architecture of main effect QTL for heading date in European winter wheat. Front. Plant Sci. 5:217. doi:10.3389/fpls.2014.00217

Zegeye, H., A. Rasheed, F. Makdis, A. Babedo, and F.C. Ogbonnaya. 2014. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. PLoS ONE 9:e105593. doi:10.1371/journal.pone.0105593

Zhang, D., G. Bai, C. Zhu, J. Yu, and B.F. Carver. 2010a. Genetic diversity, population structure and linkage disequilibrium in U.S. elite winter wheat. Plant Gen. 3:117–127. doi:10.3835/plantgenome2010.03.0004

Zhang, L., D. Liu, X. Guo, W. Yang, J. Sun, D. Wang, P. Sourdille, and A. Zhang. 2011. Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genet. 12:43. doi:10.1186/1471-2156-12-42

Zhang, L.Y., D.C. Liu, X.L. Guo, W.L. Yang, J.Z. Sun, D.W. Wang, and A. Zhang. 2010b. Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. J. Integr. Plant Biol. 52:966–1007. doi:10.1111/j.1744-7909.2010.00967.x

4 Linkage drag constrains the roots of modern wheat

Voss-Fels, K., Qian, L., Parra-Londono, S., Uptmoor, R., Frisch, M., Keeble-Gagnère, G., Appels, R., and Snowdon, R. J.

Plant, Cell & Environment (under review)

Linkage drag constrains the roots of modern wheat

Kai Voss-Fels[1], Lunwen Qian[1], Sebastian Parra-Londono[2], Ralf Uptmoor[2], Matthias Frisch[3], Gabriel Keeble-Gagnère[4], Rudi Appels[5] and Rod J. Snowdon[1*]

[1] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

[2] Department of Agronomy, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany

[3] Department of Biometry and Population Genetics, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

[4] AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport and Resources (DEDJTR), Bundoora, Victoria 3083, Australia

[5] State Agriculture Biotechnology Centre, School of Veterinary & Life Sciences, Murdoch University, Australia Export Grains Innovation Centre (AEGIC), Perth, WA, 6150, Australia

* Corresponding author:
Rod Snowdon, Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
Email: rod.snowdon@agrar.uni-giessen.de
Phone: +49 641 9937420
Fax: +49 641 9937429

Word count: 4,381

Figures: Fig. 1 (colored), Fig. 2 (colored), Fig 3 (colored)

Supporting Information (provided in additional files): Supporting Information Fig. S1-S8, Supporting Information Table S1-S8.

Abstract

Roots, the hidden half of crop plants, are essential for resource acquisition. However, knowledge about the genetic control of below-ground plant development in wheat, one of the most important small-grain crops in the world, is very limited. The molecular interactions connecting root and shoot development and growth, and thus modulating the plant's demand for water and nutrients along with its ability to access them, are largely unexplored. Here we demonstrate that linkage drag in European bread wheat, driven by strong selection for a haplotype variant controlling heading date, has eliminated a specific combination of two flanking, highly conserved, haplotype variants whose interaction confers increased root biomass. Reversing this inadvertent consequence of selection could recover root diversity that may prove essential for future food production in fluctuating environments. Highly conserved synteny to rice across this chromosome segment suggests that adaptive selection has shaped the diversity landscape of this locus across different, globally-important cereal crops. By mining wheat gene expression data we identified root-expressed genes within the region of interest that could help breeders to select positive variants adapted to specific target soil environments.

Key words: Genome-wide association study, GWAS, linkage disequilibrium, LD, linkage drag, roots, genomics, *Triticum aestivum* L. (wheat)

Introduction

Unstable environmental conditions seriously threaten the production of major crops such as wheat (*Triticum aestivum* L.), the staple food for one third of the world's population (Asseng et al. 2015). At the same time, global yield increases are unexpectedly stagnating (Ray et al. 2012; Brisson et al. 2010), a phenomenon that is frequently attributed to erosion of genetic variation in modern breeding pools (Reif et al. 2005b) and to poor root performance (White et al. 2015). The need to advance crop productivity under reduced agricultural inputs, whilst addressing adaptation to the consequences of climate change, has increased attention on roots as the primary interface for water and nutrient acquisition. Early root growth plays a critical role in nitrogen and phosphorus uptake (Waines & Ehdaie 2007), while root proliferation sustains access to nutrients and water (Den Herder et al. 2010) and can have a positive influence on grain yield (Atkinson et al. 2015). Furthermore, the physical refinement of plant stability by increasing root-soil anchorage is a crucial prerequisite for increasing the number of spikelets and grains per plant (Ren et al. 2012). Paradoxically, diminished root biomass in modern wheat cultivars has been attributed to selective breeding for grain yield under optimal water supply and fertilization (Den Herder et al. 2010; Ehdaie et al. 2003). The present scarcity of knowledge about below-ground plant development and its' genetic control is mainly due to the difficulty of measuring root features compared to above-ground plant phenotypes. Hence, further studies are essential to elucidate the molecular basis of root growth and provide reliable diagnostic molecular markers for selection (Ren et al. 2012).

On the other hand, breeding for root traits must simultaneously consider the arable value of selected genotypes, as pleiotropy or linkage drag associated with loci under selection can potentially impart negative effects on other essential traits. Negative pleitropic interactions are notoriously difficult to resolve, whereas linkage drag can potentially be overcome by

identification of rare recombinants between genes that are tightly linked in repulsion. High-resolution definition of selection hotspots underlying a joint inheritance is an important basis for targeted rejuvenation of genetic diversity and introduction of novel recombination events (Canè et al. 2014).

Pleiotropic effects of genes involved in above-ground plant development on root growth have been reported previously. For example, several *Rht* genes that played a fundamental role in maximizing the yield potential in modern wheat cultivars, through reduction of plant height, also negatively affect root proliferation (Bai et al. 2013). It is furthermore assumed that "unconscious selection" under optimal cropping conditions, in the course of intensive wheat breeding for agronomically favorable traits, led to non-optimal root systems in "Green Revolution" wheat cultivars. Three of the main green-revolution founder lines have considerably less root biomass than several landraces (Waines et al. 2007). Recent molecular studies aimed at describing the genetic background of wheat root growth used bi-parental mapping populations for QTL identification (Comas et al. 2013; Manschadi et al. 2008; Sharma et al. 2011; Bai et al. 2013), limiting the scope of the conclusions to highly specific genetic backgrounds. Genome-wide association analysis using diverse, unrelated populations can provide much more valuable insight into available diversity for breeding, but association genetic analyses of roots are extremely challenging due to the difficulties of obtaining meaningful phenotypic data for underground traits.

Here we identify genomic signatures associated with root dry mass (RDM) in a global panel of 243 genetically diverse, hexaploid wheat accessions, representing diversity from China, Europe, Australia and the United States (Voss-Fels et al. 2015), by analyzing highly significant trait associations in the context of linkage disequilibrium (LD) and population substructure. Subsequent DNA sequence and gene expression analysis identified candidate genes for both

spike and root development, indicating inadvertent co-selection of specific root variants during selection for heading behavior. We found highly synonymous regions on rice chromosome 3, suggesting that similar patterns of adaptive selection have independently conserved gene functions across this locus in different, globally-important cereal crops.

Modelling of epistasis further revealed extremely strong interactions between and among loci associated with root and spike development. Comprehensive validation experiments and high-resolution imaging confirmed the strong effect of rare genetic variants on overall root proliferation at seedling and adult plant stages, underscoring their influence on root branching, fine rooting and relative root penetration at depth. Our results establish non-pleiotropic interactions of loci associated with root and spike architecture, and enable the reversal of unintended effects on genetic variability due to linkage drag. Ongoing work based on our findings will help to functionally validate the significance of the highly conserved candidate genes described with respect to their influence on below ground development of wheat. They furthermore provide breeders the possibility for a genomics-based exchange of co-localized haplotypes to maximize root variation, while maintaining specific heading characters.

Materials and methods

Plant material and genome-wide SNP marker data

A diversity panel of 215 homozygous wheat accessions, representing species-wide diversity from China, Europe, North America and Australia, was selected from a global collection of 460 bread wheat varieties (Voss-Fels et al. 2015) based on genetic diversity analysis. The diversity panel was genotyped with the 90,000-SNP Illumina Infinium wheat genotyping array (Illumina Inc., San Diego, CA, USA) (Wang et al. 2014). The raw genotype data were filtered to retain only markers with $\leq 10\%$ missing values and minor allele frequency $\geq 5\%$, resulting in a selection of

18,855 high-quality, polymorphic SNPs for the subsequent genetic analysis. All SNP markers used for subsequent analyses were ordered according to their genetic positions in a high-resolution consensus map (Wang et al. 2014). For phenotypic validation we used additional genome-wide SNP profiles from the extended panel of 460 accessions, selecting accessions according to their haplotype patterns for *Hap-5B-RDMa* and *Hap-5B-RDMb* to construct random validation panels for the RDM I and non-RDM I haplotype groups.

Phenotypic analysis of root and shoot traits

Seedling shoot and root traits in the 215 accessions for subsequent genome-wide association analysis were assayed in greenhouse experiments (21/15°C day/night, 16h photoperiod), in three replicates of five plants each, between December 2014 and March 2015. An augmented block design with nine sub-experimental blocks was applied. Each sub-experiment contained 16 to 26 test lines plus four control lines that were tested in each of the nine sub-experiments (Supplementary Table 1). Seeds were sterilised in 6% sodium hypochlorite and sown in autoclaved soil-sand mixture (1:2 vol/vol). Seven days after sowing, plants were removed from the growth medium and transferred into sterilised 12 cm diameter plastic pots filled with autoclaved sand. Plants were continuously irrigated with 0.2% WUXAL® Super NPK hydroponic fertiliser (Wilhelm Haug GmbH & Co. KG, Düsseldorf, Germany). 35 days after sowing plants were removed from the pots and roots were carefully washed. After measuring of shoot and root length, from the base of the seedling coleoptile to the respective outermost extremity, plants were freeze-dried at -60°C for seven days before measurement of total root dry biomass for each accession.

Phenotypic validations

148 Three independent validation experiments were performed to (i) confirm the phenotypic effects
149 of the RDM I haplotype combination on root dry biomass, to (ii) confirm the biomass effects in a
150 different phenotyping system in combination with a detailed assessment of root architectural
151 factors and to (iii) analyze whether candidate genotypes with increased root systems also showed
152 superior phenotypic root values at the adult plant stage during heading. For the first validation the
153 phenotypic evaluation was performed using the same method as for the main experiment, except
154 that the ambient temperature in the greenhouse was lowered to 19/12°C day/night. In the second
155 validation experiment 24cm x 24cm plastic trays were filled with a layer of 1.6 kg sand covered
156 by a nylon mesh (20µm pore size). Trays were filled with 100 mL 0.2% WUXAL® nutrient
157 solution per kg sand and one seed per replication (with five replications per accession) was
158 germinated on top of the nylon mesh under a sealed lid for four days in the dark. Trays were
159 stacked at an angle of approximately 70° into plastic boxes until roots of the fastest-growing
160 accessions reached the extremities of the trays. After 23 days the nylon mesh was carefully
161 removed from the sand and roots were transferred to a flatbed scanner without disrupting their
162 architecture. Root images (300 dpi) generated using the scanner were analysed for 10 root
163 architectural parameters (Supporting Information Table S7) using the automated root
164 phenotyping software GIA Roots (Galkovskyi et al. 2012). Before root trait assessment, images
165 were scaled and cropped, a grey scale image was created, and a double adaptive image
166 thresholding with specific preset parameters was applied. The thresholded image sets were
167 segmented into foreground (root) and background, and pixel values were calculated scaled to
168 centimeters. In the third validation experiment, the same method was used as in the first
169 validation experiment, except for that 20 cm diameter pots were used in which four plants built
170 one replicate. At three leaf stage, the winter wheat lines were transferred to a vernalization
171 chamber for 70 days with a constant temperature of 4°C and 16h photoperiod. All other

environmental conditions were kept constant. Plants were harvested between the plant physiological stages BBCH 51-55.

Phenotypic evaluation of heading date

To obtain phenotypic heading date (HD) data, 337 of the previously described international genotypes (Voss-Fels & Snowdon 2015) were tested in a field trial at the Julius-Kühn-Institute (JKI), Braunschweig/Germany with two replicates per genotype. Seeds were sown in October 2013 and phenotypic scoring of the test plots was performed at seven different time points on May 12, 16, 20, 26 and 30, and June 3, respectively. HD was scored according to BBCH ear emergence stage 5 and flowering stage 6.

Genome wide association analysis

Genome-wide marker-trait associations were calculated from adjusted entry means for each genotype (Supporting Information Table S1), using the R package GenABEL (Aulchenko et al. 2007) and a two-step mixed linear model approach that increases detection power without increasing the empirical type I error (Stich et al. 2008). The model was adjusted for population stratification by including identity-by-state estimates for genotype pairs and a principal component adjustment that uses the first two principal components as covariates. A stringent significance cutoff value was set at $-\log10(p) = 5.57$ (5% Bonferroni threshold). To reduce the type II error rate we compared with associations captured using an arbitrary threshold of $-\log10(p) = 4$.

Linkage disequilibrium (LD) analysis and haplotype construction

Two major haplotypes for RDM were constructed by analysis of LD blocks containing SNPs showing significant marker-trait-associations (MTA) at the 5% Bonferroni threshold of $-\log_{10}$(p-value) > 5.57. Subsequently, variation for the respective RDM and HD haplotypes in the GWAS panel was assessed based on haplotype allelic state, and haplogroups were compared regarding RDM phenotypes. Haplotype networks were calculated using the program TCS v1.21(Clement, Posada & Crandall 2000).

Statistical analysis of epistatic interactions

Pairwise epistatic interactions between SNP markers located in the examined RDM haplotype region were calculated using the epistatic interaction test implemented in PLINK(Purcell et al. 2007) (v1.07, http://pngu.mgh.harvard.edu/purcell/plink/) with a linear model as:

$$y_i = \mu + a_1 SNP1_i + a_2 SNP2_i + a_3 SNP1xSNP2_i + e_i \ ,$$

where y_i represents the RDM phenotype of genotype i, a_1 and a_2 represent the effects of the two respective SNP markers, and a_3 is the interaction effect (i.e. epistatic interaction) of the two considered SNPs. To correct the multiple comparisons a stringent Bonferroni threshold (α=0.05) was used.

Bioinformatics analysis

The markers identified as significant in the GWAS were used to define the region of interest in a high-density genetic consensus map of wheat (Wang et al. 2014) constructed using the same 90k SNP Illumina genotying array. Using BLASTN, these markers were aligned to the wheat chromosome 5B pseudomolecule downloaded from Ensemblplants 30. This defined the region of interest in the wheat chromosome 5BL physical sequence. Ortholog data for wheat, rice and sorghum genes was obtained from Ensemblplants using the Biomart interface and used to project

219 the region of interest from wheat chromosome 5B into rice chromosome 3, and from rice
220 chromosome 3 into sorghum chromosome 1, respectively. This information was subsequently
221 organised into tab-delimited files and used as input into a custom D3.js visualisation to produce
222 an SVG image of the alignment. This file was imported into the software package Inkscape
223 (https://www.inkscape.org) to produce a high-resolution image for publication. Expression
224 profiles were inspected using the web-based tool Tritigate (Nystrom-Persson et al., manuscript in
225 preparation), loaded with gene expression data from the wheat expression database
226 http://www.wheat-expression.com (Borrill et al. 2016). The main dataset used to study root
227 expression was that reported in (Pingault et al. 2015).
228
229 **Results**
230 **Eco-geographical wheat gene pools show high variation for basic seedling traits**
231 Initially, we assayed root dry mass (RDM), leaf dry mass (LDM), root length (RL), shoot length
232 (SL) and root to shoot ratio (R/S) of 215 wheat accessions (Supporting Information Table S1) in
233 nine greenhouse experiments 35 days after sowing. Applying an augmented design to obtain
234 comparable adjusted mean phenotype values from three replicates of five plants (Materials and
235 Methods) we found that the two subpopulations representing Chinese wheat gene pools showed
236 distinctly higher phenotypic means than European wheat subpopulations, especially for RDM and
237 RL (Supporting Information Fig. S1A, Supporting Information Table S2). Although the larger
238 European subgroup EU2 exhibited similar SL to Chinese germplasm, the LDM in EU2 was
239 considerably lower than in CHN1 and CHN2. Significant positive correlations could be found
240 among almost all traits except SL, which was not correlated with R/S. The strongest correlations
241 were found between RDM and R/S ($r=0.91$, $p<0.001$), RDM and LDM ($r=0.54$, $p<0.001$) and
242 LDM and SL ($r=0.57$, $p<0.001$) (Supporting Information Fig. S1B), respectively. Interestingly,

243 RL had a higher correlation with R/S than LDM or SL, and was correlated to LDM in a similar
244 manner like RDM.
245
246 **Root-associated SNP haplotypes directly flank a conserved heading-date QTL**
247 Genome-wide marker-trait association studies (GWAS) and local LD analyses were performed
248 using single-nucleotide polymorphism (SNP) data, generated for the diversity panel using the
249 Illumina 90k SNP Infinium array (Wang et al. 2014). Correcting for population substructure,
250 genome-wide SNP-trait association analysis revealed two adjacent loci, at 137.1 cM and 143.5
251 cM on chromosome 5B, with highly significant associations to RDM that exceeded a stringent
252 Bonferroni threshold of –log10(p-value) > 5.57 (Fig. 1a, Supporting Information Table S3). None
253 of the other traits investigated exhibited marker-trait-associations with p-values below the
254 significance threshold. However, five SNPs on 2B and 5A with –log10(p-value) > 4 had a strong
255 negative effect on RL, while one SNP on 6D was strongly associated to SL. Elucidation of local
256 LD revealed that the RDM-associated SNPs lie in two neighboring, highly conserved haplotype
257 blocks which we designated *Hap-5B-RDMa* and *Hap-5B-RDMb*. Nine SNPs with significant
258 associations to RDM were assigned to *Hap-5B-RDMa* and six SNPs were assigned to *Hap-5B-*
259 *RDMb*, respectively (Supporting Information Table S4,S5). Mean RDM values were significantly
260 higher in accessions that combined the positive haplotype variants *Hap-5B-RDMa-2* and *Hap-5B-*
261 *RDMb-3* (subsequently referred to as RDM_I accessions) than in accessions carrying only one or
262 neither of these variants (Fig. 2b). Positive epistatic interaction among these two loci was
263 confirmed by the epistatic interaction test implemented in the whole genome association analysis
264 toolset PLINK (v1.07, http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al. 2007) (Fig. 3).
265 Haplotype network analysis and comparison with population substructure revealed carriers of the
266 RDM_I haplotype combination exclusively among Chinese wheat accessions. Furthermore,

RDM_I accessions were genetically distinct from European accessions representing the most frequent haplotype variants, *Hap-5B-RDMa-1* and *Hap-5B-RDMb-1* (Supporting Information Fig. S2, Fig. 2). A notable block with highly conserved LD, comprising 46 SNPs, separates the root-haplotype blocks *Hap-5B-RDMa* and *Hap-5B-RDMb* (Fig. 1*B*). These 46 SNPs show extremely strong LD ($r^2 \geq 0.8$) in all but the Chinese subpopulation CHN2, and 14 of them are known to have strong associations to heading date (HD) (Zanke et al. 2014). A further study recently identified two QTL in this region at 139.6 and 144 cM, with a high influence on inflorescence architecture and paired spikelet development (Boden et al. 2015), hence we assumed strong prior directional selection on this genomic region and designated it *Hap-5B-HD*. A total of 38 unique haplotype variants were observed for *Hap-5B-HD* in the haplotype network analysis, whereas only seven and nine haplotype variants were identified for *Hap-5B-RDMa* and *Hap-5B-RDMb*, respectively (Supporting Information Tables S4-S6, Fig. 2a). Comparing the frequency and distributions of the *Hap-5B-HD* variants in the four RDM groups showed that the most prominent variants were represented in each subgroup (Supporting Information Fig. S3).

Independent validations confirm haplotype effects on seedling and adult plant root growth

To validate the phenotypic effects of the *Hap-5B-RDMa-2/Hap-5B-RDMb-3* haplotype combination in RDM_I accessions, we screened genome-wide SNP data in 245 additional, non-phenotyped hexaploid wheat accessions from an international wheat collection (Voss-Fels et al. 2015). We found 8 lines in total that carry the rare RDM_I haplotype combination (*Hap-5B-HDa-2* and *Hap-5B-HDb-3*) and compared their root dry mass with 13 randomly selected non-RDM_IV accessions that carry neither of the two favorable RDM haplotype variants (Supporting Information Table S1). Highly significant differences in root biomass ($p<0.0001$) were confirmed between these two groups (Supporting Information Fig. S4). For a second, independent

validation, and to obtain additional information on the root system architecture underlying the differences in root mass, we used the automated root phenotyping software GIA Roots (Galkovskyi et al. 2012) to analyse 10 parameters describing root morphology, after 23 days of *in vitro* rhizotron growth, in 15 RDM_I and 9 RDM_IV accessions representing both the original test panel and the validation panel (Materials and Methods). RDM_I accessions showed significant differences in multiple parameters describing the horizontal and vertical distribution of the roots and in the size and density of the root system, compared to RDM_IV haplotype combinations (Fig. 2c,d; Supporting Information Fig. S5A-G). A third validation experiment was performed to confirm the influence of the RDM_I haplotype combination on root development in adult plants. For this reason, we compared R/S of 40 genotypes representing RDM_I and non-RDM_I accessions at heading (BBCG 51-59). These included 20 spring (7 RDM_I, 13 non-RDM genotypes) and 20 winter (8 RDM_I, 12 non-RDM_I genotypes) wheat lines. RDM_I genotypes showed significantly higher R/S measurements than non-RDM genotypes in both spring ($p=0.021$) and winter (0.0033) lines (Supporting Information Fig. S6). To verify differences in spike development in the four different RDM groups, we investigated heading date at seven different time points, during a field trial in 2014, in 337 international genotypes (Voss-Fels et al. 2015) that represented all four RDM subgroups (Supporting Information Fig. S7). While accessions from RDM_I produced spikes significantly earlier, with over 80% of genotypes flowering by May 22, only 50% of the lines from RDM_IV (which includes mostly European elite varieties) flowered by June 3. Among the RDM_I genotypes only the breeding line W-WW331, derived from a cross between European and Chinese lines, showed a similar heading behavior to the RDM_IV lines.

Haplotype blocks on 5B are synonymous to rice chromosome 3 and harbor genes highly expressed in root and spike tissues

The SNP markers described in Fig. 1 define the region of interest spanning the haplotype blocks *Hap-5B-RDMa*, *Hap-5B-HD* and *Hap-5B-RDMb* on wheat chromosome 5B. Based on BLAST positions of the SNP flanking sequences, this region corresponds to the genomic sequence of the wheat chromosome 5B pseudomolecule from position 249,920,295bp (marker Tdurum_contig47627_442) to 260,041,085bp (marker RAC875_c24226_1356), according to the Ensemblplants 30 pseudomolecules for *Triticum aestivum*. A total of 213 protein-coding genes are defined in this region by the Ensemblplants 30 annotation. Using the Ensemblplants "Genomic alignments" tool, we found this region to be syntenic to the terminal end of rice chromosome 3. Due to the inherent difficulty of defining a genomic region via genetic markers ordered by linkage, we redefined the region of interest as follows. Firstly, the orthologous relationships between wheat and rice genes were used to project the relevant segment of wheat chromosome 5B onto rice chromosome 3. Subsequently, the genes and orthologous relationships in the corresponding chromosome segment from rice were used to project back to a slightly expanded region of chromosome 5B. This added 29 genes to the 5BL region that are not present on the current version of the wheat chromosome 5B pseudomolecule. We also inspected the syntenic relationship between this region of rice chromosome 3 and sorghum chromosome 1. The relevant 1.3 Mb region of rice chromosome 3 (from 33.5 Mbp to 34.8 Mbp) defines 215 genes in the International Rice Genome Sequence Project (IRGSP) genome annotation. Of these genes, 159 have orthologous relationships to genes in wheat; 1247 of these are on wheat chromosome 5BL and 11520 (934%) show non-zero expression in root tissue, based on RNA-Seq expression data from http://www.wheat-expression.com (9 genes were not contained in the database) (Borrill et al. 2016). Since both the rice genome sequence (gene order) and annotation is currently more

advanced than that of wheat, we focused attention on the genes in the wheat 5BL region with syntenic relationships to rice, using the rice functional annotation as a proxy for the wheat gene, and for which the rice ortholog also had an ortholog in the equivalent region in sorghum. The list of genes and orthology information, with the starting point being the region as defined in rice, is summarised in Supporting Information Table S8. The markers mapping to the relevant region of the wheat 5B pseudomolecule, together with syntenic relationships between the wheat, rice and sorghum genes in this region, are shown in Fig. 1a.

Discussion

Rare genetic variants on chromosome 5B drive root proliferation in Chinese wheat

Climatic changes and resulting fluctuating environmental conditions require root system improvement of modern wheat varieties. However, strong selective breeding under optimal fertilizer and irrigation inputs appear to have diminished root mass (Waines et al. 2007; Den Herder et al. 2010). Furthermore, there is increasing concern that breeding has drastically narrowed overall genetic diversity in elite germplasm to an extent that insufficient variation is still available for effective selection (Reif et al. 2005a). Here we discovered two extremely rare haplotype variants, on chromosome 5B in exotic Chinese wheat lines, whose interaction associates with a drastic increase in overall root proliferation. The profound genetic distance of these variants to European elite lines makes them an ideal source to enhance the genetic predisposition to produce increased root systems.

Breakage of linkage drag among loci linked in repulsion depends on identification of rare recombinants among desirable allelic variants. Using high-resolution SNP genotype profiles, we confirmed the existence of desirable recombination events between European and Chinese wheat genotypes. These recombinants combine desirable root and flowering haplotypes in European-

362 adapted wheat, reuniting European *Hap-5B-HD* variants with the rare RDM variants *Hap-5B-*
363 *RDMa-2* and *Hap-5B-RDMb-3* that interact to enhance root biomass. Marker-assisted
364 backcrossing to introgress this novel haplotype combination into further elite European winter
365 wheats will elucidate the genomic effect of these variants in multiple genetic backgrounds of
366 high-yielding varieties. The high correlations we observed of RDM to R/S and RL to LDM
367 suggest that the roots play an important role in optimizing the ratio of above-ground to and
368 below-ground biomass, a critical aspect of adaptation to increasing extreme environmental
369 conditions.
370
371 **Selection for heading behavior has diminished genetic variation for root growth in**
372 **European wheat**
373 Different wheat gene pools show profound genomic signatures of directional selection, with
374 highly conserved LD surrounding important loci for eco-geographic adaptation and agricultural
375 performance (Voss-Fels et al. 2015). The high LD we found around the extended chromosomal
376 block *Hap-5B-HD*, a locus that has been ascribed a high effect on spike development and heading
377 date, indicate strong prior directional selection for this genomic region. The strong representation
378 of the HD-associated haplotype variants *Hap-5B-HD-1*, *Hap-5B-HD-2, Hap-5B-HD-3* and *Hap-*
379 *5B-HD-4* in every RDM group (Supporting Information Fig. S3) indicates that the Chinese
380 RDM_I haplotype variants share ancestry with HD variants in RDM group IV, encompassing
381 elite European wheat varieties. The genotype clustering based on genetic distances further
382 suggests that RDM_I accessions have not been used in the establishment of European cultivars
383 (Supporting Information Fig. S2). The very close proximity of the flanking root-associated
384 haplotype blocks *Hap-5B-RDMa* and *Hap-5B-RDMb* has apparently caused co-selection of
385 specific RDM haplotype variants due to linkage, which diminished genetic variation for root

386 growth in recent elite European wheat germplasm. This knowledge provides breeders with a
387 simple opportunity to quickly reinstate root diversity into the European bread wheat gene pool,
388 which might prove essential in times of unexpected climatic fluctuations.
389
390 **Haplotype blocks contain candidate genes highly expressed in root and spike tissues**
391 Within the group of genes expressed in roots, three were of particular interest to root morphology
392 due to their role in carbohydrate/fibre synthesis: The genes Traes_5BL_CFCBFDA99.2
393 (annotated as a cellulose synthase), Traes_5BL_1133E46E7.1 (an expansin) and
394 Traes_5BL_1512240F3.1 (an endo-beta-mannanase) showed only low expression in roots. On
395 the other hand, their putative involvement in generation of signal molecules means that only
396 small transcript quantities are normally necessary to impart regulatory effects (Zhao et al. 2013).
397 Endo-beta-mannanases and expansins represent two classes of genes that are of particular interest
398 to root phenotypes, since both form a part of the gene network involved in cell expansion
399 (Hrmova et al. 2006; Zhao et al. 2013; Gonzalez-Calle et al. 2015; He et al. 2015). The evidence
400 in the present manuscript that they are associated with root QTL in wheat suggests they are good
401 targets for more detailed analysis. The endo-beta-mannanase has a striking pattern of expression
402 (low expression in roots early in development and more highly expressed in the grain late in
403 development, Supporting Information Fig. S8) which may provide a useful indicator for defining
404 other components of gene network involved in root development. This apparent role in root
405 developmental processes supports an involvement of Traes_5BL_1512240F3.1 in root
406 development, and may provide a useful indicator for defining other interacting gene network
407 components.
408
409

Acknowledgments

R.J.S and K.V.-F. acknowledge funding from the German Federal Ministry of Food and Agriculture (BMEL), grant number FNR-22408212. We also thank Stefan Kontowski (W. von Borries-Eckendorf GmbH & Co. KG.) and Bertrand Schuiling (Wiersum Plantbreeding BV) for providing large parts of the wheat collection. Further, we thank Andreas Welke, Stavros Tzigos, Annette Plank, Birgit Keiner, Liane Renno, Nelly Weis and Sebastian Brinker for their excellent technical assistance. We would also like to thank Bernd Rodemann (Julius Kühn-Institut, Braunschweig) for providing the field data on heading date.

References

Asseng S., Ewert F., Martre P. et al. (2015) Rising temperatures reduce global wheat production. *Nature Clim. Change* **5,** 143–147.

Atkinson J.A., Wingen L.U., Griffiths M. et al. (2015) Phenotyping pipeline reveals major seedling root growth QTL in hexaploid wheat. *Journal of experimental botany* **66,** 2283–2292.

Aulchenko Y.S., Ripke S., Isaacs A. & van Duijn, Cornelia M (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics (Oxford, England)* **23,** 1294–1296.

Bai C., Liang Y. & Hawkesford M.J. (2013) Identification of QTLs associated with seedling root traits and their correlation with plant height in wheat. *Journal of experimental botany* **64,** 1745–1753.

Boden S.A., Cavanagh C., Cullis B.R., Ramm K., Greenwood J., Jean Finnegan E., Trevaskis B. & Swain S.M. (2015) Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants* **1,** 14016.

Borrill P., Ramirez-Gonzalez R. & Uauy C. (2016) expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant physiology* **170,** 2172–2186.

Brisson N., Gate P., Gouache D., Charmet G., Oury F.-X. & Huard F. (2010) Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Research* **119,** 201–212.

Canè M.A., Maccaferri M., Nazemi G., Salvi S., Francia R., Colalongo C. & Tuberosa R. (2014) Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Molecular breeding : new strategies in plant improvement* **34,** 1629–1645.

Clement M., Posada D. & Crandall K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular ecology* **9,** 1657–1659.

Comas L.H., Becker S.R., Cruz, Von Mark V, Byrne P.F. & Dierig D.A. (2013) Root traits contributing to plant productivity under drought. *Frontiers in plant science* **4,** 442.

Den Herder G., van Isterdael G., Beeckman T. & Smet I. de (2010) The roots of a new green revolution. *Trends in plant science* **15,** 600–607.

Ehdaie B., Whitkus R.W. & Waines J.G. (2003) Root biomass, water-use efficiency and performance of wheat-rye translocations of chromosomes 1 and 2 in spring wheat pavon. *Crop Science,* 710–717.

Galkovskyi T., Mileyko Y., Bucksch A. et al. (2012) GiA Roots: software for the high throughput analysis of plant root system architecture. *BMC Plant Biology* **12,** 116.

Gonzalez-Calle V., Barrero-Sicilia C., Carbonero P. & Iglesias-Fernandez R. (2015) Mannans and endo-beta-mannanases (MAN) in Brachypodium distachyon: expression profiling and possible role of the BdMAN genes during coleorhiza-limited seed germination. *Journal of experimental botany* **66,** 3753–3764.

He X., Zeng J., Cao F., Ahmed I.M., Zhang G., Vincze E. & Wu F. (2015) HvEXPB7, a novel beta-expansin gene revealed by the root hair transcriptome of Tibetan wild barley, improves root hair growth under drought stress. *Journal of experimental botany* **66,** 7405–7419.

Hrmova M., Burton R.A., Biely P., Lahnstein J. & Fincher G.B. (2006) Hydrolysis of (1,4)-beta-D-mannans in barley (Hordeum vulgare L.) is mediated by the concerted action of (1,4)-beta-D-mannan endohydrolase and beta-D-mannosidase. *The Biochemical journal* **399,** 77–90.

Manschadi A.M., Hammer G.L., Christopher J.T. & deVoil P. (2008) Genotypic variation in seedling root architectural traits and implications for drought adaptation in wheat (Triticum aestivum L.). *Plant and Soil* **303,** 115–129.

Pingault L., Choulet F., Alberti A., Glover N., Wincker P., Feuillet C. & Paux E. (2015) Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome biology* **16,** 29.

Purcell S., Neale B., Todd-Brown K. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81,** 559–575.

Ray D.K., Ramankutty N., Mueller N.D., West P.C. & Foley J.A. (2012) Recent patterns of crop yield growth and stagnation. *Nature communications* **3,** 1293.

Reif J.C., Zhang P., Dreisigacker S., Warburton M.L., van Ginkel M., Hoisington D., Bohn M. & Melchinger A.E. (2005a) Wheat genetic diversity trends during domestication and breeding. *Theor Appl Genet* **110,** 859–864.

Reif J.C., Zhang P., Dreisigacker S., Warburton M.L., van Ginkel M., Hoisington D., Bohn M. & Melchinger A.E. (2005b) Wheat genetic diversity trends during domestication and breeding. *Theor Appl Genet* **110,** 859–864.

Ren Y., He X., Liu D., Li J., Zhao X., Li B., Tong Y., Zhang A. & Li Z. (2012) Major quantitative trait loci for seminal root morphology of wheat seedlings. *Molecular Breeding* **30,** 139–148.

Sharma S., Xu S., Ehdaie B., Hoops A., Close T.J., Lukaszewski A.J. & Waines J.G. (2011) Dissection of QTL effects for root traits using a chromosome arm-specific mapping population in bread wheat. *Theor Appl Genet* **122,** 759–769.

Stich B., Möhring J., Piepho H.-P., Heckenberger M., Buckler E.S. & Melchinger A.E. (2008) Comparison of mixed-model approaches for association mapping. *Genetics* **178,** 1745–1754.

Voss-Fels K., Frisch M., Qian L., Kontowski S., Friedt W., Gottwald S. & Snowdon R.J. (2015) Subgenomic Diversity Patterns Caused by Directional Selection in Bread Wheat Gene Pools. *The Plant Genome* **8,** 0.

Voss-Fels K. & Snowdon R.J. (2015) Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant biotechnology journal,* n/a.

Waines J.G. & Ehdaie B. (2007) Domestication and crop physiology: roots of green-revolution wheat. *Annals of botany* **100,** 991–998.

Wang S., Wong D., Forrest K. et al. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant biotechnology journal* **12,** 787–796.
White C.A., Sylvester-Bradley R. & Berry P.M. (2015) Root length densities of UK wheat and oilseed rape crops with implications for water capture and yield. *Journal of experimental botany* **66,** 2293–2303.
Zanke C., Ling J., Plieske J. et al. (2014) Genetic architecture of main effect QTL for heading date in European winter wheat. *Frontiers in plant science* **5,** 217.
Zhao Y., Song D., Sun J. & Li L. (2013) Populus endo-beta-mannanase PtrMAN6 plays a role in coordinating cell wall remodeling with suppression of secondary wall thickening through generation of oligosaccharide signals. *The Plant journal : for cell and molecular biology* **74,** 473–485.

Figure legends

Figure 1. Genome-wide association mapping results for root dry mass on Chromosome 5B, synteny map for orthologous genes and subgroup specific pairwise linkage disequilibrium (LD) patterns in this region. (a) Each dot represents the $-\log_{10}$(p-value) of a SNP marker-trait-association. The dashed line shows the 5% Bonferroni threshold ($-\log_{10}$(p-value) = 5.57). The colored heat maps represent pairwise LD as r^2 between marker pairs in the genetic subgroups that have been identified in population structure analysis (Supporting Information Fig. S2). Synteny map shows SNP alignment from the genetic map (Wang et al. 2014) onto the physical map. Orthologous genes across rice and sorghum are indicated by lines. (b) Pairwise LD measures as r^2 between markers in the 5B chromosomal block from 135.5 to 144.4 cM. LD was calculated for all genotypes (N=215) and for the identified genetic subgroups China I (N=11), China II (N=46), Mixed (N=54), Europe I (N=25) and Europe II (N=79) separately. Brown and blue arrows indicate the position of root (*Hap-5B-RDMa* and *–b*) and spike (*HD-5B-HD*) associated haplotype SNPs (Materials and Methods, Supporting Information Table S4-S6).

Figure 2. Effects of combined root dry mass (RDM) haplotype variants on root dry mass. (a) Haplotype networks of the two RDM-related haplotypes *Hap-5B-RDMa* and *-b*. The pie charts represent haplotype variants, their sizes are proportional to the number of genotypes that carry the respective variant and the colors indicate the gene pool affiliation. The networks give estimates on the genealogies of the haplotype variant sequences. All haplotype variants are differentiated from the nearest variant by one-nucleotide change in the haplotype sequence. Grey circles indicate additional probable sequence altering steps between two haplotype variations. The two haplotype variants that are associated to high RDM are highlighted. (b) The four groups

represent genotypes that differ in the presence (+) / absence (-) of the favorable haplotype variants *Hap-5B-RDMa-2* and *Hap-5B-RDMb-3* and the boxplots show their corresponding RDM phenotypes. RDM_I= + / +, RDM_II= + / -, RDM_III= - / +, RDM_IV= - / -. ANOVA significance thresholds for pairwise comparisons of RDM groups II-IV with RDM_I indicated as * (p=0.05), ** (p=0.01), *** (p=0.001) and - (no significance). (c) Root images of six representatives of RDM_I and (d) RDM_IV genotypes at 23 days after sowing. Additional numbers -1 to -5 indicate the replicate number in the scanning experiment.

Figure 3. Heat map of epistatic interactions between single nucleotide polymorphism (SNP) markers in the root dry mass (RDM) related candidate region on chromosome 5B. Region spanning from 136.5 to 144.5 cM. Colors indicate pairwise SNP x SNP interactions from blue (≙ significant negative effect on RDM) to red (≙ significant positive effect on RDM).

(Supporting Information Figures are provided in a separate file)

Supporting Information Fig. S1. Phenotypic variation among genetic subgroups and trait correlation matrix. (a) Mean values of the whole population and the five genetic subgroups among the 215 genotypes in the GWAS panel. (b) Pearson correlations among all traits for RDM = Rood dry mass, LDM = Leaf dry mass, RL = Root length, SL = Shoot length, R/S = Root to shoot ratio with significance thresholds * (p=0.05), ** (p=0.01), *** (p=0.001) and - (no significance); Scatterplots include LOESS fitting curves.

Supporting Information Fig. S2. Principal component analysis of the tested wheat lines. Calculations are based on modified Roger's distances for 20,283 polymorphic single nucleotide polymorphism (SNP) markers (minor allele frequency ≥ 5%) and 215 genotypes. Colored ellipses represent the origins of the tested lines.

Supporting Information Fig. S3. Haplotype network of the heading date (HD) related haplotype Hap-5B-HD. (a) The pie charts represent haplotype variants, their sizes are proportional to the number of genotypes that carry the respective variant and the colors indicate the gene pool affiliation. The networks give estimates on the genealogies of the haplotype variant sequences. All haplotype variants are differentiated from the nearest variant by one-nucleotide change in the haplotype sequence. Grey circles indicate additional probable sequence altering

steps between two haplotype variations. (b) The colored bars represent haplotype variation frequencies for Hap-5B-HD in the respective RDM groups I (n=7), II (n=11), III (n=4), and IV (n=193).

Supporting Information Fig. S4. Phenotypic evaluation of the root dry mass (RDM) related haplotypes on RDM among 21 wheat lines. RDM was measured 35 days after growing with three replicates per genotype consisting of five plants per replication. RDM_I genotypes carry both favorable RDM haplotype variants Hap-5B-RDMa-2/ Hap-5B-RDMb-3 while RDM_IV genotypes carry none of the positive variants. ANOVA significance threshold is *** (p=0.001).

Supporting Information Fig. S5. Phenotypic evaluation of the root dry mass (RDM) related haplotypes on ten visual root traits among 24 wheat lines analysed with GiA roots software. Ten different traits were measured using the visual root analysis software GiA roots for seedling roots 23 days after sowing of which seven exceeded highly significant differences between RDM_I (n=15) and IV (n=9) genotypes (ANOVA significance threshold: *** p<0.001, - n.s.). (a) MNR= Maximum number of roots. (b) MeNR= Median number of roots. (c) NA= Network area. (d) NP= Network perimeter. (e) NSA= Network surface area. (f) NLD= Network length distribution. (g) NV= Network volume. (h) ARW= Average root width. (i) NB= Network bushiness. (j) SRL= Specific root length. Detailed trait description in Supporting Information Table S7.

Supporting Information Fig. S6. Phenotypic evaluation of the root dry mass (RDM) related haplotypes on adult plants' RDM. (a) Root-to-shoot ratio (RS) was measured at BBCH stage 51-55 with three replicates per genotype consisting of four plants each. A total of 40 genotypes were analysed (20 genotypes per growth type group). RDM_I genotypes (seven for spring, eight for winter group) carry both favorable RDM haplotype variants Hap-5B-RDMa-2/ Hap-5B-RDMb-3 while non-RDM_I genotypes do not carry this allelic combination. ANOVA significance threshold is * (p=0.05) and ** (p=0.01).

Supporting Information Fig. S7. Frequency of spike emergence and flowering time among 337 genotypes at seven different scoring time points. 183 genotypes of the 337 have been tested in at least one panel of this experiment. Plants were tested in field trials at

Braunschweig/Germany in 2014 in two replicate plots per genotype. Number of genotypes tested: RDM_I = 7, RDM_II = 6, RDM_III = 5, RDM_IV = 319. Heading = BBCH 51-59, Flowering = BBCH 61-69).

Supporting Information Fig. S8. Expression levels for selected genes. TPM (transcripts per million) expression levels downloaded from www.wheat-expression.com (Borrill et al. 2016) for Traes_5BL_CFCBFDA99.2, a cellulose synthase, Traes_5BL_1512240F3.1, an endo-beta-mannanase, and Traes_5BL_1133E46E7.1, an expansin in (a) wheat variety 'Chinese Spring' across different plant tissues and (b) in root tissue across all available varieties. The TPM mean value and the standard error are shown. dpa = days post anthesis. N1DT1A, N1AT1B, N1AT1D, N1BT1A, N1DT1B, N1BT1D, N5BT5A, N5AT5B, N5AT5D, N5DT5A, N5DT5A, N5DT5A and N5BT5D are Nulli-Tetra lines of 'Chinese Spring' that miss a given chromosome. Two additional copies of another chromosome compensate for the absence of the respective chromosome; e.g. N5BT5D has four copies of 5D that compensate for the absence of 5B.

Fig. 1

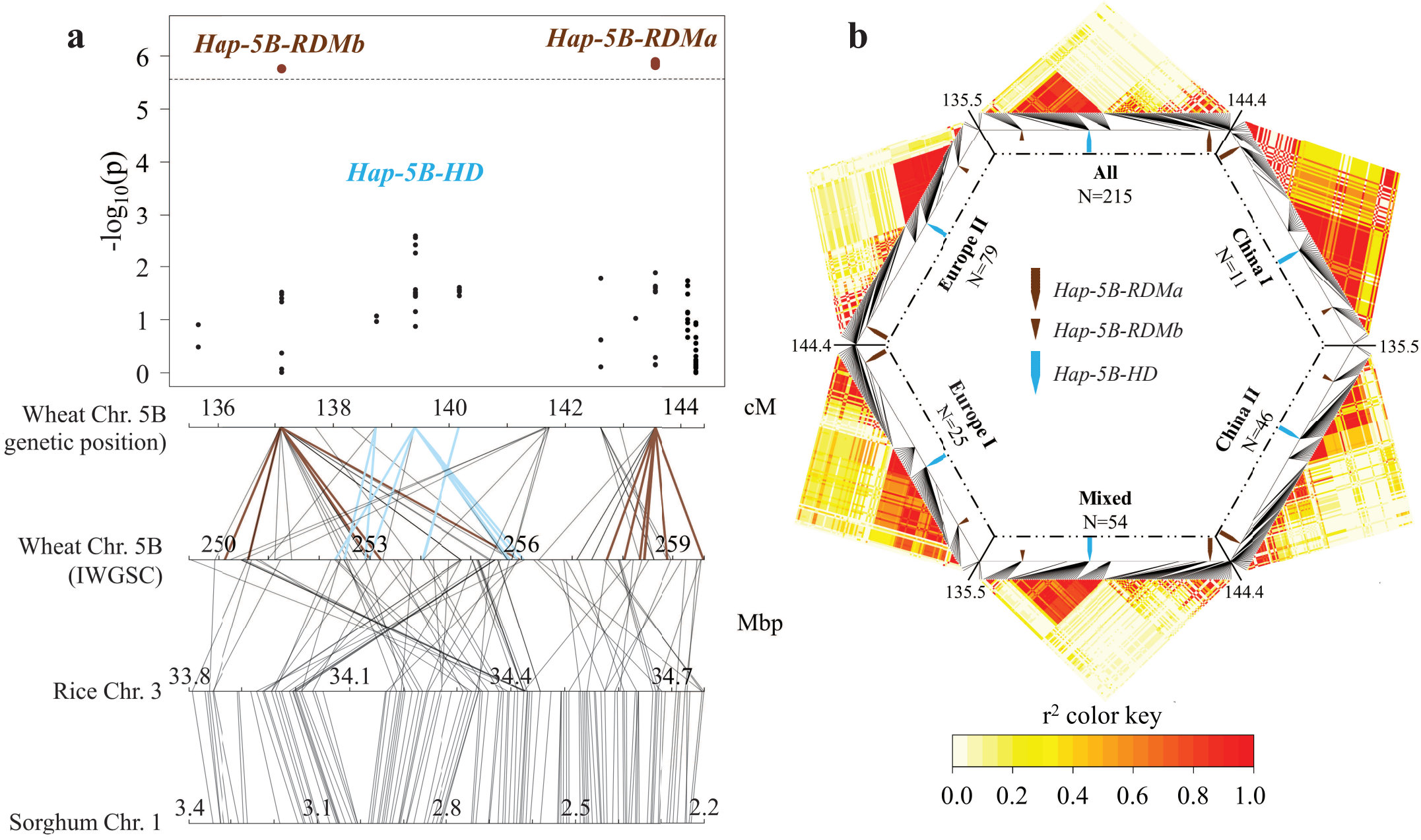

a
Hap-5B-RDMb
Hap-5B-RDMa
Hap-5B-HD
-log10(p)
0
1
2
3
4
5
6
Wheat Chr. 5B genetic position)
136
138
140
142
144
cM
Wheat Chr. 5B (IWGSC)
250
253
256
259
Mbp
Rice Chr. 3
33.8
34.1
34.4
34.7
Sorghum Chr. 1
3.4
3.1
2.8
2.5
2.2
b
135.5
144.4
All
N=215
China I
N=11
China II
N=46
Mixed
N=54
Europe I
N=25
Europe II
N=79
Hap-5B-RDMa
Hap-5B-RDMb
Hap-5B-HD
r2 color key
0.0
0.2
0.4
0.6
0.8
1.0

Fig. 2

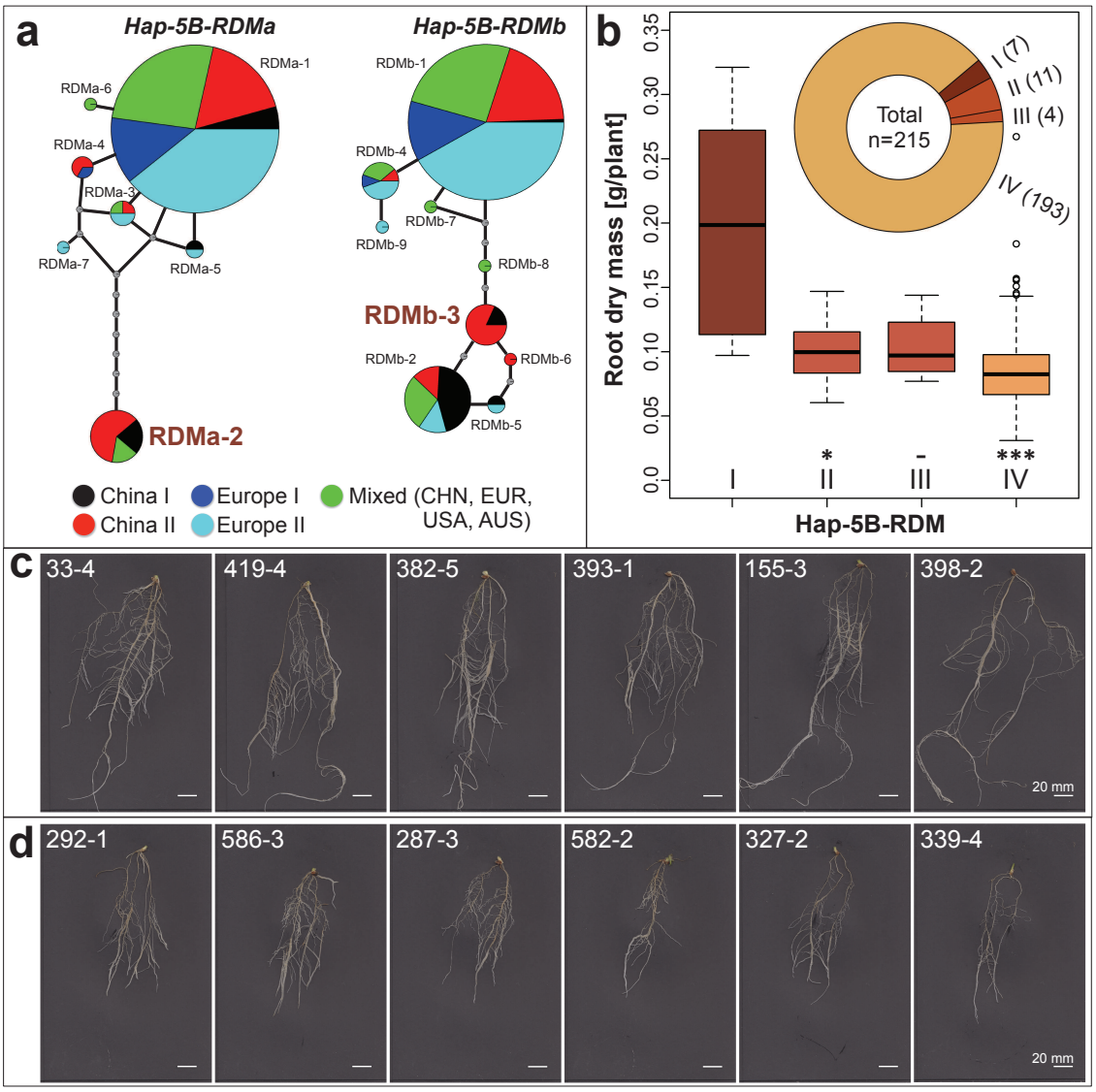
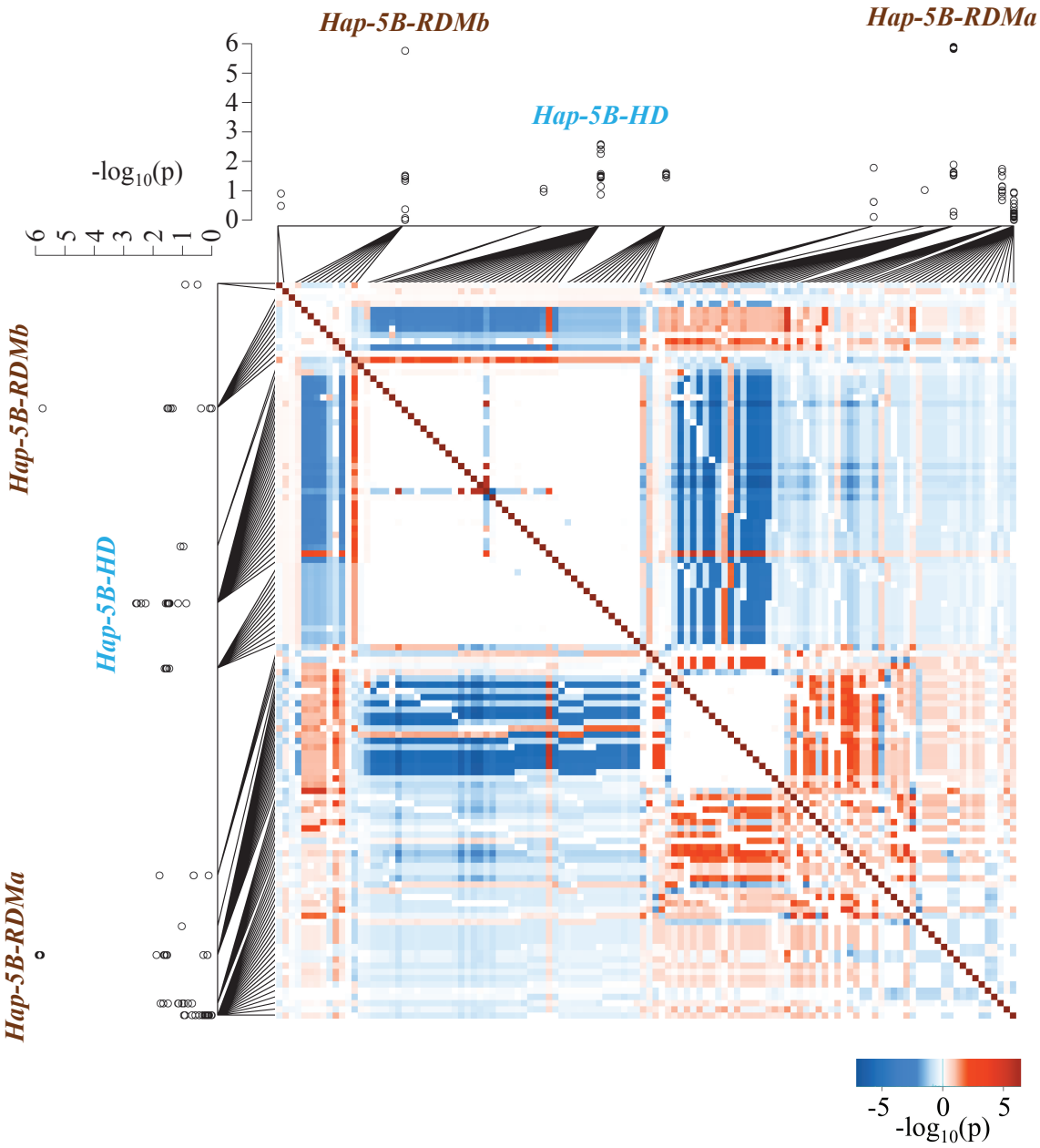

a
Hap-5B-RDMa
Hap-5B-RDMb
RDMa-1
RDMa-6
RDMa-4
RDMa-3
RDMa-7
RDMa-5
RDMa-2
RDMb-1
RDMb-4
RDMb-7
RDMb-9
RDMb-8
RDMb-3
RDMb-2
RDMb-6
RDMb-5
China I
China II
Europe I
Europe II
Mixed (CHN, EUR, USA, AUS)
b
Root dry mass [g/plant]
0.0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Total n=215
I (7)
II (11)
III (4)
IV (193)
*
-
***
I
II
III
IV
Hap-5B-RDM
c
33-4
419-4
382-5
393-1
155-3
398-2
20 mm
d
292-1
586-3
287-3
582-2
327-2
339-4
20 mm

Fig. 3


Hap-5B-RDMb
Hap-5B-RDMa
Hap-5B-HD
-log10(p)
6
5
4
3
2
1
0
Hap-5B-RDMb
Hap-5B-HD
Hap-5B-RDMa
-5
0
5
-log10(p)

5 Discussion

5.1 Evaluating and increasing genetic diversity in wheat

The projected sharp increase in demand for wheat within the next decades, coupled with the limited area available for agricultural production, necessitate a substantial improvement of production systems, and particularly the sustained development of improved, high-yielding cultivars (Shiferaw et al., 2013). However, an extensive exchange of genetic material between modern breeding programs and strong selection have narrowed the elite germplasm pool, to an extent that breeders cannot create sufficient variation from which they can continuously select improved genotypes (Reif et al., 2005). A key future approach to break this so-called "yield barrier" is to include exotic germplasm in modern breeding programs. Such variants exhibit a multitude of novel auspicious variation and account for the largest part of wheat genetic resources. Despite this value, the vast majority of the available genetic diversity has only been used in modern wheat improvement to a very limited extent (Longin and Reif, 2014; Mondal et al., 2016). One reason is that this approach has been inefficient in practice due to the poor performance of crosses with wild relatives in target growing environments, necessitating costly and time consuming breeding steps (Huang and Han, 2014; Lopes et al., 2015). With the invention of next-generation sequencing technologies, novel and highly efficient genomic tools for the establishment of precise genome-wide genotyping profiles have been developed in all important crops. Such technologies are nowadays in standard use in wheat genomic research (Chapter 2). Different commercial genotyping services are available that provide thousands or millions of data points per genotyping experiment, and platforms that are able to detect polymorphisms on a single-nucleotide basis between genotypes are today widespread (Bassi et al., 2016). Such data can be invaluable in wheat improvement strategies, particularly as a resource to assess genetic variability in cultivars, to identify optimal parental combinations for the creation of segregating populations with maximal genetic diversity, and to perform marker-

based targeted introgressions of candidate genes by crossing elite lines with high yield potential to exotic lines that carry favorable novel alleles. This enables marker-assisted selection of genotypes that carry the desired introgression while maximizing the genetic background from the elite parent (Mohammadi and Prasanna, 2003). The work described in this thesis assessed various population genetic measures in a diverse population of 460 international hexaploid wheat lines, using the Illumina 90K SNP Infinium genotyping array (Voss-Fels et al., 2015; Chapter 3). The accessions in the diversity panel originate from different parts of China, Europe, Australia and the United States. They include high-yielding winter wheat varieties, most of them with a strong European background, breeding lines from running commercial breeding programs, along with exotic landraces from China with spring growth habit. The population was assembled by commercial wheat breeders in order to introduce novel genetic diversity into European gene pools and to introgress new sources of biotic resistances.

A principal component analysis (PCA)-based k-means clustering approach identified five major subgroups, which generally matched available information about the origin and growth habits of the lines. In-depth population structure analysis of the five different subgroups further revealed stratification in each of the groups, with differing average levels of genetic diversity among the corresponding lines. These findings can guide breeders to select parents from crossings between or within gene pools, to increase genetic variation in breeding approaches. The study thus provides a valuable basis for the combination of genetic materials that carry high levels of phenotypic and genotypic variation, but simultaneously have similar adaptation characteristics in order to maximize diversity for selection. Interestingly, seven lines designated to originate from Chinese germplasm collections clustered closely with European elite lines, while three lines from the panel that were assumed to be independent genotypes were genetically 100% identical with

the Chinese landrace Wangshubai, indicating possible errors during the panel establishment. These examples demonstrate the power of this genotyping technology in detecting true genetic relationships between lines, which might otherwise be confounded by incorrect recordings and traditional pedigree information from conventional breeding approaches.

Analysis of average genetic diversity and linkage disequilibrium (LD) in the three subgenomes of hexaploid wheat revealed the highest of loss of variation, accompanied by allelic fixation, in the D-subgenome. This result, consistent with previous findings (Benson et al., 2012; Chao et al., 2010; Chen et al., 2012; Nielsen et al., 2014; Wang et al., 2014), can mainly be attributed to the most recent allopolyploidization event, with the D-subgenome being the last subgenome that completed hexaploid wheat. This hybridization is considered to have caused a significant allopolyploidization bottleneck (Marcussen et al., 2014). The low level of genetic diversity in the D-subgenome, even among extremely diverse gene pools (e.g. China vs. Europe), underlines the pressing need to rejuvenate novel variation in this subgenome, e.g. by establishing synthetic wheat lines through crossings between ancient wheat relatives (Henry and Nevo, 2014; Jia et al., 2013).

While many different population genetic studies have been published that characterize diverse wheat populations with various molecular marker systems, investigations that pinpoint specific chromosomal regions with strong signatures of directional selection (also known as ‘selective sweeps’) are rare. The present study distinguished several hotspots that have experienced drastic losses of allelic variation, most likely caused by directional selection, on twelve chromosomes covering all three subgenomes. Of special interest was the region between 59 and 72 cM on chromosome 1B, which harbors almost 50% of all identified SNP candidates that were found to be under statistically significant selection. Detailed analysis of pairwise LD in this chromosomal

block revealed high levels of genetic fixation in contrasting allelic states between the two major gene pools from China and Europe, accompanied by a very slow LD decay in each of the five genetic subgroups. According to numerous findings from the literature, there are several QTL, meta-QTL and candidate genes in this block that influence agronomically important traits, such as yield, plant height or baking quality. Marker-guided crossing of suitable parents could help to facilitate recombination in this conserved chromosomal block and to develop novel variation for selection.

5.2 The molecular connection between spike and root development in wheat

The improvement of wheat yields in order to increase the overall global production is particularly important in marginal growing regions with unfavorable environmental conditions and very low outputs per area (Tester and Langridge, 2010). Improving the resilience of modern wheat cultivars towards various environmental stresses as a consequence of climate change is a major goal. Especially water shortage and droughts are expected to become increasingly frequent and severe in the near future (Seneviratne et al., 2012). In this context, roots have gained increasing attention, as they represent the primary interface for nutrient and water uptake of the plant (Den Herder et al., 2010). However, wheat researchers have almost exclusively focused on above-ground plant characteristics and roots have largely been ignored, due to the notorious difficulty in obtaining effective phenotyping measurements underground.

While there are a number of wheat root QTL mapping studies using bi-parental mapping populations, such approaches are inherently limited to the allelic diversity of the crossing parents, and the relatively low degree of recombination (Bai et al., 2013; Comas et al., 2013; Korte and Farlow, 2013; Manschadi et al., 2008; Sharma et al., 2011). Against this background, a GWAS

was performed for basic growth characteristics in a panel of 243 diverse international wheat lines (Chapter 4). Genotypes were selected based on population genetic analyses (Chapter 3) in order to maintain the broad diversity of the initial international population of 460 lines, while reducing the number of genotypes that needed to be tested in comprehensive greenhouse screenings. In total, this collection contained five different genetic subgroups (Voss-Fels et al., 2015), comprising two from China, two from Europe and one mixed group that contained genotypes from various origins. At first, root dry mass (RDM), leaf dry mass (LDM), root length (RL), shoot length (SL) and root to shoot ratio (R/S) were assayed in 215 wheat accessions, 35 days after sowing. This revealed that the two Chinese genotype clusters, mainly containing landraces, had distinctly higher phenotypic means, especially for RDM and RL, compared to the two European subpopulations which include registered varieties and germplasm from ongoing breeding programs. This is consistent Waines and Ehdaie (2007), who report a drastic decrease of roots in elite germplasm compared to several landraces. This phenomenon could easily be explained by modern breeding approaches, in which the best genotypes are often selected under optimal water and nutrient input environments (Ehdaie et al., 2003). Analysis of correlations between the measured traits revealed that R/S, a trait that can convey an adaptive advantage for drought at pre-anthesis (Manschadi et al., 2008), is highly correlated with RDM ($r=0.91$, $p<0.001$), while the correlations with LDM are comparatively low ($r=0.19$, $p<0.01$). This indicates that root growth is an important adjustment screw in the context of R/S, rather than above-ground plant growth.

The GWAS, using almost 19,000 polymorphic SNPs, identified two highly associated QTL on chromosome 5B that have not been described previously. It is assumed that this candidate region was previously undetected due insufficient allelic variation in earlier mapping studies. The two

root-associated loci, found within an interval of ~7 cM, exhibit very high measures of LD, leading to a high level of allelic conservation in each of the five subgroups. To further increase the allelic variation for the two identified QTL, and to increase the phenotypic resolution in specific genetic groups, haplotypes were constructed based on LD between the significant SNPs from the GWAS and their flanking markers (Jiang et al., 2015). In total, seven and nine haplotype variants were identified for the root-associated haplotypes *Hap-5B-RDMa* and *–b*, respectively, whereas 38 different haplotype alleles were identified for a completely conserved intermediate block that has previously been associated with spike development (Boden et al., 2015; Zanke et al., 2014). This block was subsequently referred to as *Hap-5B-HD*. Given the prevalent high LD in the whole identified block, very strong directional selection for this region was assumed, in accordance to the fact that spike development and heading date, which also affect flowering time, have been under very strong natural and artificial selection in the course of local adaptation and crop improvement by breeding (Cockram et al., 2007).

By assigning the genotypes to haplotype groups according to their allelic state and comparing their phenotypic measures, it was found that the seven lines which carry a combination of *Hap-5B-RDMa-2* and *–b-3* have significantly higher root biomass compared to other lines and were subsequently assigned to subgroup RDM_I. Epistatic interaction among these two loci was confirmed by the epistatic interaction test implemented in the whole genome association analysis toolset PLINK (v1.07, http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007). Comparison with population substructure revealed carriers of the RDM_I haplotype combination exclusively among Chinese wheat accessions. Furthermore, RDM_I accessions were genetically distinct from European accessions representing the most frequent haplotype variants, *Hap-5B-*

RDMa-1 and *Hap-5B-RDMb-1*, suggesting that RDM_I accessions have not been used in breeding of European wheat.

To validate the phenotypic effects of the *Hap-5B-RDMa-2/Hap-5B-RDMb-3* haplotype combination in RDM_I accessions, genome-wide SNP data were analyzed in 245 additional, non-phenotyped hexaploid wheat accessions from an international wheat collection (Voss-Fels et al., 2015). Assays of root dry mass in a selection of 8 further selected RDM_I accessions (carrying *Hap-5B-HDa-2* and *Hap-5B-HDb-3*) from the validation panel, compared to 13 randomly-selected accessions from the validation panel that carry neither of the two favorable RDM haplotype variants, confirmed highly significant differences in root biomass ($p<0.0001$) between these two groups. For a second, independent validation, and to obtain additional information on the root system architecture underlying the differences in root mass, the automated root phenotyping software GiA Roots (Galkovskyi et al., 2012) was used to analyse 10 parameters describing root morphology, after 23 days of *in vitro* rhizotron growth, in 15 RDM_I and 9 RDM_IV (random genotypes that carry any other alleles for *Hap-5B-RDMa* and *Hap-5B-RDMb*) accessions, representing both the original test panel and the validation panel. RDM_I accessions showed significant differences in multiple parameters describing the horizontal and vertical distribution of the roots, and in the size and density of the root system, compared to RDM_IV haplotype combinations. A third validation experiment was performed to confirm the influence of the RDM_I haplotype combination on root development in adult plants. Again, highly significant increases were seen in RDM_I genotypes compared to non-RDM genotypes, in both spring and winter wheat.

The most prevalent heading-associated haplotype variants, *Hap-5B-HD-1*, *Hap-5B-HD-2, Hap-5B-HD-3* and *Hap-5B-HD-4*, are strongly represented in every RDM group, indicating that the

Chinese RDM_I haplotype variants share ancestry with HD variants within the RDM group IV that have been selected in European wheat breeding. The very close proximity of the flanking *Hap-5B-RDMa* and *Hap-5B-RDMb* haplotype blocks has apparently caused co-selection of specific RDM haplotype variants due to linkage rather than pleiotropy. Existing recombinants between European and Chinese wheat genotypes, such as breeding line W-WW311, combine desirable *Hap-5B-HD* variants with the rare RDM variants *Hap-5B-RDMa-2* and *Hap-5B-RDMb-3*, thus fulfilling the flowering requirements for local adaptation. The production of further introgression lines that carry this novel diversity will elucidate the genomic effect of these variants in multiple genetic backgrounds of high-yielding varieties. This knowledge provides breeders with a simple opportunity to quickly reinstate root diversity into the European bread wheat gene pool, while maintaining desired heading characteristics.

Subsequent bioinformatics analysis and sequential alignment identified adjacent candidate genes for both, spike and root development, in the candidate region, further indicating inadvertent co-selection of specific root variants during breeding for heading date. The proximity of candidate genes on the physical map, along with their described functions, suggests the existence of potential epistatic interactions. For example, the identified region contains a gene model that is annotated as an endo-beta mannase-encoding gene. This well-studied gene family, recognized through conserved Pfam domains within the clan GH-A, encodes endohydrolase enzymes known to modify cell wall structure in tomato (Bewley, 1997) or coffee (Marraccini et al., 2001), for example. Subsequent functional validation of promising gene candidates will elucidate the direct phenotypic effects and might lay the foundation for future biotechnology-based approaches that enable a targeted gene customization to ultimately improve root systems in modern cultivars.

5.3 Conclusions and future prospects

This thesis describes the utilization possibilities of modern, high-throughput genotyping technologies for future wheat research and their applications for breeding. These automated tools, commercially available for all important crop plants, enable the rapid, cheap and ultra-fast establishment of whole-genome marker profiles for large plant populations, at an unprecedented resolution, with millions of genome-wide data points simultaneously. One of their major applications is the detailed genomic characterization of germplasm, a crucial prerequisite for the establishment of optimal crossing schemes in breeding programs. Within the framework of this dissertation, population structure, genetic diversity and linkage disequilibrium (LD) was investigated in a large and diverse population of international wheat lines at the sub-genome and chromosome level, using a novel high-throughput genotyping platform. Identifying major genetic subpopulations, along with varying levels of diversity and allelic fixation among and between groups, this work revealed major hotspots in the complex hexaploid wheat genome that are under very strong directional selection between different gene pools. Breeders can directly use this information to reinstate genetic variation in these target regions in order to broaden overall diversity as the prerequisite for selection.

Further, this work identified a novel and yet completely unknown chromosomal region that harbors genomic signatures associated with root growth, a trait that has been vastly ignored in wheat research to date but has never been more important in the context of adapting new cultivars to extreme climate changes. Population genetic analysis reveals that extremely rare Chinese QTL variants boost overall root proliferation, and that these identified QTL are tightly linked to genes that affect spike development and heading date. Collectively, these results suggest that an inadvertent co-selection of root-affecting genes during selection for spike development has

narrowed genetic variation for below-ground plant development in elite lines. This knowledge provides breeders with a simple opportunity to quickly reinstate root diversity into the European bread wheat gene pool while maintaining desired heading characteristics.

6 Summary

Wheat (*Triticum aestivum* L.), the staple foot for one third of the world population, is one of the three most important crop plants worldwide. In 2014, 730 M t of wheat seeds were produced on an area of 220 M ha, with a global average yield of 3.6 t/ha. To meet the projected demand of 70 % more food by the middle of this century, wheat yields must be almost doubled within the next few decades. However, trends indicate current stagnation of yield levels in the major wheat growing regions of Europe, North America, Asia and Australia. This is mainly attributed to a narrow genetic diversity in elite germplasm pools, causing poor cultivar adaption to an increasing frequency of extreme climatic fluctuations.

Increasing genetic variation in breeding programs is a key approach to overcome yield stagnation, but it requires efficient tools that enable breeders to characterize large populations of genotypes at a maximum resolution and low cost. Due to the high complexity of the allohexaploid wheat genome ($2n = 6x = 42$ chromosomes, genomic code AABBDD), which arose from three independent allopolyploidization events, deep genome analysis is notoriously difficult. Rapid genotyping platforms that can simultaneously acquire millions of genomic data points per genotype are today available for all important crop plants. These can be used to obtain population genetic parameters for the analysis of prevalent population stratification, genetic diversity or linkage disequilibrium (LD). This information is highly valuable for breeders, in order to optimize parental combinations for the creation of segregation populations with maximal genetic diversity, and to support marker-based targeted introgression of candidate genes into elite varieties, by crossing lines with high yield potentials and exotic lines that carry favorable novel alleles.

In this context, this thesis describes the detailed population genetic characterization of a diverse international population of 460 wheat lines, representing global diversity from China, Europe, The United States and Australia, using the 90K single-nucleotide-polymorphism (SNP) Infinium genotyping platform. After identifying five major genetic subgroups that correspond to the ecogeographic origin of the lines, detailed measures of genetic diversity and LD were obtained at the subgenome and chromosome level. Finally, target regions were identified on all three subgenomes that are under very strong directional selection between the characterized gene pools. These "selective sweeps" harbor several candidate genes and QTL for agronomically important traits, such as yield, plant height or seed quality, and are thus ideal targets for a precise reinstatement of genetic variation using high-resolution marker information. This can help breeders to increase overall diversity as a key prerequisite for effective selection in breeding programs.

The need to adapt modern wheat cultivars to the increasing frequency of extreme weather events has considerably raised attention on roots. However, knowledge on the genetic basis of below-ground plant development is very limited, yet powerful diagnostic markers for a genomics-based selection of improved root systems are strongly required. This dissertation describes the first genome-wide association study (GWAS) on root characteristics in hexaploid wheat to date, using a diverse population of almost 250 wheat lines in combination with high-resolution, genome-wide SNP array marker information. After identifying two QTL on chromosome 5B that are highly associated with root biomass, haplotype network analysis attributed the highest genetic effect to two rare Chinese QTL alleles whose positive epistatic interaction boosts overall root proliferation. Population genetic analysis revealed high levels of LD in this target region, which was also found to harbor candidate genes affecting spike development and heading date. Given

the low level of allelic diversity of this candidate region, an indication of strong directional selection, the frequency and distribution of root and spike associated haplotype variants suggests an inadvertent co-selection of specific root-related variants in the course of selection for spike development. This could explain the low genetic diversity for these root QTL in European elite material. These findings provide breeders a means to reverse this unintended consequence of linkage drag and increase genetic variation for root growth, and lay a foundation for the subsequent functional characterization of candidate genes that affect root development in wheat.

7 Zusammenfassung

Weizen (*Triticum aestivum* L.), das Grundnahrungsmittel für ein Drittel der Weltbevölkerung, ist eine der drei wichtigsten Ackerkulturpflanzen weltweit. Im Jahr 2014 wurden 730 Mio. t Weizenkörner auf einer Fläche von 220 Mio. ha bei einem Durchschnittsertrag von 3.6 t/ha geerntet. Um die vorhergesagte Nachfrage nach 70 % mehr Lebensmitteln bis zur Mitte dieses Jahrhunderts zu decken, muss der weltweite Weizenertrag in den nächsten Jahrzehnten etwa verdoppelt werden. Gleichzeitig zeigen globale Anbautrends eine Ertragsstagnation in den Hauptbewirtschaftungsregionen in Europa, Nordamerika, Asien und Australien. Das kann hauptsächlich der sehr stark eingeschränkten genetischen Diversität in Elite-Genpools zugeschrieben werden, welche mitunter dazu führt, dass Sorten schlechte Anpassungseigenschaften an immer häufiger werdende klimatische Veränderungen zeigen.

Um der Ertragsstagnation entgegenzuwirken ist es daher ein wesentlicher Ansatzpunkt, die genetische Variation in Züchtungsprogrammen zu erhöhen. Dafür benötigt es jedoch effiziente Methoden, die es den Züchtern ermöglichen, große Populationen hochauflösend und kostengünstig genomisch zu charakterisieren. Aufgrund der hohen Komplexität des allohexaploiden Weizengenoms (2n = 6x = 42 Chromosomen, genomischer Code AABBDD), welches in drei unabhängigen Allopolyploidisierungsereignissen entstand, ist die tiefgehende Genomanalyse stark erschwert. Hocheffiziente Genotypisierungsplattformen, welche in kürzester Zeit Millionen von genomischen Datenpunkten generieren können, sind heutzutage für alle wichtigen Ackerkulturpflanzen verfügbar. Diese Daten können dazu genutzt werden, populationsgenetische Parameter für die Untersuchung der vorherrschenden Populationsstruktur, der genetischen Diversität oder des Gametenphasenungleichgewichts (engl.: „Linkage Disequilibrium“; LD) zu berechnen. Für Züchter stellen diese Informationen wertvolle

Grundlagen dar, um die parentalen Kombinationen zur Erstellung von Züchtungspopulationen mit maximierter genetischer Diversität zu optimieren, oder um gezielte Marker-gestützte Einkreuzungen von Genen in Elitesorten, realisiert durch die Kreuzung von Linien mit hohem Ertragspotential und beispielsweise exotischen Material mit neuen, bevorzugten Allelen, zu unterstützen.

Die vorliegende Doktorarbeit beschreibt in diesem Zusammenhang die detaillierte, populationsgenetische Charakterisierung einer diversen, aus 460 internationalen Weizenlinien bestehenden Population, welche globale Diversität aus China, Europa, den Vereinigten Staaten von Amerika und Australien repräsentiert, mit der 90K Single-Nukleotid-Polymorphismus (SNP) Infinium Genotypisierungsplattform. Nach der Identifizierung von fünf genetischen Hauptgruppen, welche mit dem ökogeographischen Ursprung der Weizenlinien korrespondieren, wurden umfangreiche Daten zur genetischen Diversität und LD auf Subgenom- und Chromosomenebene erhoben. Zudem konnten genomische Zielregionen auf allen drei Subgenomen ausfindig gemacht werden, welche Anzeichen für starke gerichtete Selektion zwischen den ermittelten Genpools aufweisen. Diese sogenannten „Selective Sweeps“ decken verschiedene Gene und „Quantitative Trait Loci“ (QTL) ab, welche agronomisch wichtige Merkmale, wie Ertrag, Wuchshöhe oder Kornqualität beeinflussen und daher ideale Ausgangspunkte für die Wiederherstellung genetischer Variation mittels hochauflösender Markerdaten darstellen. Dies kann Züchtern dabei helfen, die Diversität, welche die Grundlage für eine effektive Selektion in Zuchtprogrammen darstellt, gezielt zu erhöhen.

Die Notwendigkeit, moderne Sorten an die stetig zunehmenden extremen Wetterbedingungen anzupassen, hat die Aufmerksamkeit für Wurzeln stark erhöht. Dennoch ist das aktuelle Wissen über die genetische Basis des unterirdischen Pflanzenwachstums stark begrenzt, auch wenn ein

starker Bedarf nach effektiven diagnostischen Markern für eine Genom-gestützte Selektion verbesserter Wurzelsysteme besteht. Die vorliegende Dissertation beschreibt die bislang erste genomweite Assoziationsstudie (GWAS) zu Wurzelmerkmalen in einer Population von fast 250 diversen, hexaploidem Weizenlinien in Kombination mit hochauflösenden, genomweiten SNP Array Markerdaten. Nach Identifizierung zweier QTL auf Chromosom 5B, welche stark mit Wurzelbiomasse assoziiert sind, konnte mittels Haplotypnetzwerkanalyse der stärkste phänotypische Effekt zwei seltenen, chinesischen QTL-Allelen zugeschrieben werden, dessen positive epistatische Interaktion das Gesamtwurzelwachstum stark erhöht. Populationsgenetische Analysen zeigten hohe LD-Werte in dieser Region, welche zudem Kandidatengene für Ährenentwicklung und Blühzeitpunkt enthält. Vor dem Hintergrund der sehr geringen allelischen Diversität in diesem chromosomalen Bereich, was ein Anzeichen für starke, gerichtete Selektion darstellt, suggerierte die Frequenz und Verteilung der Wurzel- und Ähren-assoziierten Haplotypvarianten eine zufällige Ko-Selektion spezieller Wurzel-bezogener Allele während der Selektion für Ährenentwicklung. Das könnte die niedrige genetische Diversität für die gefundenen Wurzel-QTL in europäischem Elitematerial erklären. Die vorliegenden Ergebnisse können von Züchtern genutzt werden, um diese ungewünschte Konsequenz der Genkopplung rückgängig zu machen und die genetische Variation für Wurzelwachstum zu erhöhen. Sie dienen darüber hinaus als Grundlage für zukünftige Arbeiten zur funktionellen Charakterisierung von Kandidatengenen, welche das Wurzelwachstum in Weizen steuern.

8 References

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., Hayden, M. J., Howes, N., Sharp, P., Vaughan, P., Rathmell, B., Huttner, E. and Kilian, A. (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *TAG Theoretical and Applied Genetics* **113**, 1409–1420.

Anderson, J. A., Stack, R. W., Liu, S., Waldron, B. L., Fjeld, A. D., Coyne, C., Moreno-Sevilla, B., Fetch, J. M., Song, Q. J., Cregan, P. B. and Frohberg, R. C. (2001) DNA markers for Fusarium head blight resistance QTLs in two wheat populations. *TAG Theoretical and Applied Genetics* **102**, 1164–1168.

Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., Foresman, B. J. and Kolb, F. L. (2016a) Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Wheat using Genotyping-by-Sequencing. *The Plant Genome* **9**, 0.

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., Brown-Guedira, G., Dong, Y., Foresman, B. J. and Kolb, F. L. (2016b) Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (Triticum aestivum L.). *Mol Breeding* **36**.

Asseng, S., Ewert, F., Martre, P., Rotter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., Sanctis, G. de, Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Muller, C., Naresh Kumar, S., Nendel, C., O/'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z. and Zhu, Y. (2015) Rising temperatures reduce global wheat production. *Nature Clim. Change* **5**, 143–147.

Atkinson, J. A., Wingen, L. U., Griffiths, M., Pound, M. P., Gaju, O., Foulkes, M. J., Le Gouis, J., Griffiths, S., Bennett, M. J., King, J. and Wells, D. M. (2015) Phenotyping pipeline reveals major seedling root growth QTL in hexaploid wheat. *Journal of experimental botany* **66**, 2283–2292.

Baenziger, P. S., Bakhsh, A., Lorenz, A. and Walia, H. (2014) Bridging Conventional Breeding and Genomics for A More Sustainable Wheat Production. In: *Genomics of Plant Genetic*

Resources (Tuberosa, R., Graner, A. and Frison, E., eds), pp. 185–209. Dordrecht: Sprer Netherlands.

Bai, C., Liang, Y. and Hawkesford, M. J. (2013) Identification of QTLs associated with seedling root traits and their correlation with plant height in wheat. *Journal of experimental botany* **64**, 1745–1753.

Balat, M., Balat, H. and Öz, C. (2008) Progress in bioethanol processing. *Progress in Energy and Combustion Science* **34**, 551–573.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R. and Crossa, J. (2016) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). *Plant science an international journal of experimental plant biology* **242**, 23–36.

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., Fritz, A. K. and Poland, J. A. (2016) Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *The Plant Genome* **9**, 0.

Battisti, D. S. and Naylor, R. L. (2009) Historical warnings of future food insecurity with unprecedented seasonal heat. *Science (New York, N.Y.)* **323**, 240–244.

Benson, J., Brown-Guedira, G., Paul Murphy, J. and Sneller, C. (2012) Population Structure, Linkage Disequilibrium, and Genetic Diversity in Soft Winter Wheat Enriched for Fusarium Head Blight Resistance. *The Plant Genome Journal* **5**, 71.

Bewley, J. D. (1997) Breaking down the walls — a role for endo-β-mannanase in release from seed dormancy? *Trends in plant science* **2**, 464–469.

Bhullar, N. K., Street, K., Mackay, M., Yahiaoui, N. and Keller, B. (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9519–9524.

Boden, S. A., Cavanagh, C., Cullis, B. R., Ramm, K., Greenwood, J., Jean Finnegan, E., Trevaskis, B. and Swain, S. M. (2015) Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature plants* **1**, 14016.

Bonman, J. M., Babiker, E. M., Cuesta-Marcos, A., Esvelt-Klos, K., Brown-Guedira, G., Chao, S., See, D., Chen, J., Akhunov, E., Zhang, J., Bockelman, H. E. and Gordon, T. C. (2015) Genetic Diversity among Wheat Accessions from the USDA National Small Grains Collection. *Crop Science* **55**, 1243.

Borrell, A. K., Mullet, J. E., George-Jaeggli, B., van Oosterom, E. J., Hammer, G. L., Klein, P. E. and Jordan, D. R. (2014) Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *Journal of experimental botany* **65**, 6251–6263.

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X. and Huard, F. (2010) Why are wheat yields stagnating in Europe?: A comprehensive data analysis for France. *Field Crops Research* **119**, 201–212.

Cane, M. A., Maccaferri, M., Nazemi, G., Salvi, S., Francia, R., Colalongo, C. and Tuberosa, R. (2014) Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Molecular breeding new strategies in plant improvement* **34**, 1629–1645.

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., da Silva, M. L., Bockelman, H., Talbert, L., Anderson, J. A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P. L., Dubcovsky, J., Morell, M. K., Sorrells, M. E., Hayden, M. J. and Akhunov, E. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8057–8062.

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S. P., Matnyazov, R., Clark, D. R., Talbert, L. E., Anderson, J. A., Dreisigacker, S., Glover, K., Chen, J., Campbell, K., Bruckner, P. L., Rudd, J. C., Haley, S., Carver, B. F., Perry, S., Sorrells, M. E. and Akhunov, E. D. (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.). *BMC genomics* **11**, 727.

Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y., Tang, X., Zhou, F., Deng, X. W. and Zhang, Q. (2014) A high-density SNP genotyping array for rice biology and molecular breeding. *Molecular plant* **7**, 541–553.

Chen, X., Min, D., Yasir, T. A. and Hu, Y.-G. (2012) Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PloS one* **7**, e44510.

Cockram, J., Jones, H., Leigh, F. J., O'Sullivan, D., Powell, W., Laurie, D. A. and Greenland, A. J. (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *Journal of experimental botany* **58**, 1231–1244.

Collard, B. C. Y. and Mackill, D. J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **363**, 557–572.

Comas, L. H., Becker, S. R., Cruz, V. M. V., Byrne, P. F. and Dierig, D. A. (2013) Root traits contributing to plant productivity under drought. *Frontiers in plant science* **4**, 442.

Den Herder, G., van Isterdael, G., Beeckman, T. and Smet, I. de (2010) The roots of a new green revolution. *Trends in plant science* **15**, 600–607.

Deng, W., Casao, M. C., Wang, P., Sato, K., Hayes, P. M., Finnegan, E. J. and Trevaskis, B. (2015) Direct links between the vernalization response and other key traits of cereal crops. *Nature communications* **6**, 5882.

Edwards, D., Batley, J. and Snowdon, R. J. (2013) Accessing complex crop genomes with next-generation sequencing. *TAG Theoretical and Applied Genetics* **126**, 1–11.

Ehdaie, B., Whitkus, R. W. and Waines, J. G. (2003) Root Biomass, Water-Use Efficiency, and Performance of Wheat–Rye Translocations of Chromosomes 1 and 2 in Spring Bread Wheat 'Pavon'. *Crop Science* **43**, 710.

Ellis, J. G., Lagudah, E. S., Spielmeyer, W. and Dodds, P. N. (2014) The past, present and future of breeding rust resistant wheat. *Frontiers in plant science* **5**, 641.

Fahlgren, N., Gehan, M. A. and Baxter, I. (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current opinion in plant biology* **24**, 93–99.

Fleury, D., Jefferies, S., Kuchel, H. and Langridge, P. (2010) Genetic and genomic tools to improve drought tolerance in wheat. *Journal of experimental botany* **61**, 3211–3222.

Fu, Y.-B., Peterson, G. W., Yu, J.-K., Gao, L., Jia, J. and Richards, K. W. (2006) Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. *TAG Theoretical and Applied Genetics* **112**, 1239–1247.

G20 (2011). *Ministerial Declaration: Action plan on food price volatility and agriculture*. Paris.

Galkovskyi, T., Mileyko, Y., Bucksch, A., Moore, B., Symonova, O., Price, C. A., Topp, C. N., Iyer-Pascuzzi, A. S., Zurek, P. R., Fang, S., Harer, J., Benfey, P. N. and Weitz, J. S. (2012) GiA Roots: software for the high throughput analysis of plant root system architecture. *BMC Plant Biol* **12**, 116.

Ganal, M. W., Durstewitz, G., Polley, A., Berard, A., Buckler, E. S., Charcosset, A., Clarke, J. D., Graner, E.-M., Hansen, M., Joets, J., Le Paslier, M.-C., McMullen, M. D., Montalent, P., Rose, M., Schon, C.-C., Sun, Q., Walter, H., Martin, O. C. and Falque, M. (2011) A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PloS one* **6**, e28334.

Ganal, M. W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H. and Durstewitz, G. (2012) Large SNP arrays for genotyping in crop plants. *J Biosci* **37**, 821–828.

Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-l., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science (New York, N.Y.)* **296**, 92–100.

Grainger, M. (2010) World Summit on Food Security (UN FAO, Rome, 16–18 November 2009). *Development in Practice* **20**, 740–742.

Gupta, P. K., Mir, R. R., Mohan, A. and Kumar, J. (2008) Wheat genomics: present status and future prospects. *International journal of plant genomics* **2008**, 896451.

Hedden, P. (2003) The genes of the Green Revolution. *Trends in Genetics* **19**, 5–9.

Henry, R. J. and Nevo, E. (2014) Exploring natural selection to guide breeding for agriculture. *Plant biotechnology journal* **12**, 655–662.

Huang, X. and Han, B. (2014) Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology* **65**, 531–551.

Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., Appels, R., Pfeifer, M., Tao, Y., Zhang, X., Jing, R., Zhang, C., Ma, Y., Gao, L., Gao, C., Spannagl, M., Mayer, K. F. X., Li, D., Pan, S., Zheng, F., Hu, Q., Xia, X., Li, J., Liang, Q., Chen, J., Wicker, T., Gou, C., Kuang, H., He, G., Luo, Y., Keller, B., Xia, Q., Lu, P., Wang, J., Zou, H., Zhang, R., Xu, J., Gao, J., Middleton, C., Quan, Z., Liu, G., Wang, J., Yang, H., Liu, X., He, Z., Mao, L. and Wang, J. (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95.

Jiang, Y., Jiang, Q., Hao, C., Hou, J., Wang, L., Zhang, H., Zhang, S., Chen, X. and Zhang, X. (2015) A yield-associated gene TaCWI, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis. *TAG Theoretical and Applied Genetics* **128**, 131–143.

Joukhadar, R., El-Bouhssini, M., Jighly, A. and Ogbonnaya, F. C. (2013) Genome-wide association mapping for five major pest resistances in wheat. *Mol Breeding* **32**, 943–960.

Koebner, R. M. and Summers, R. W. (2003) 21st century wheat breeding: Plot selection or plate detection? *Trends in Biotechnology* **21**, 59–63.

Korte, A. and Farlow, A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* **9**, 29.

Lesk, C., Rowhani, P. and Ramankutty, N. (2016) Influence of extreme weather disasters on global crop production. *Nature* **529**, 84–87.

Li, S., Jia, J., Wei, X., Zhang, X., Li, L., Chen, H., Fan, Y., Sun, H., Zhao, X., Lei, T., Xu, Y., Jiang, F., Wang, H. and Li, L. (2007) A intervarietal genetic map and QTL analysis for yield traits in wheat. *Mol Breeding* **20**, 167–178.

Liu, G., Zhao, Y., Gowda, M., Longin, C. F. H., Reif, J. C. and Mette, M. F. (2016) Predicting Hybrid Performances for Quality Traits through Genomic-Assisted Approaches in Central European Wheat. *PloS one* **11**, e0158635.

Longin, C. F. H., Muhleisen, J., Maurer, H. P., Zhang, H., Gowda, M. and Reif, J. C. (2012) Hybrid breeding in autogamous cereals. *TAG Theoretical and Applied Genetics* **125**, 1087–1096.

Longin, C. F. H. and Reif, J. C. (2014) Redesigning the exploitation of wheat genetic resources. *Trends in plant science* **19**, 631–636.

Longin, C. F. H. and Würschum, T. (2016) Back to the Future - Tapping into Ancient Grains for Food Diversity. *Trends in plant science*.

Lopes, M. S., Dreisigacker, S., Pena, R. J., Sukumaran, S. and Reynolds, M. P. (2015) Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. *TAG Theoretical and Applied Genetics* **128**, 453–464.

Manifesto, M. M., Schlatter, A. R., Hopp, H. E., Suárez, E. Y. and Dubcovsky, J. (2001) Quantitative Evaluation of Genetic Diversity in Wheat Germplasm Using Molecular Markers. *Crop Science* **41**, 682.

Manschadi, A. M., Hammer, G. L., Christopher, J. T. and deVoil, P. (2008) Genotypic variation in seedling root architectural traits and implications for drought adaptation in wheat (Triticum aestivum L.). *Plant Soil* **303**, 115–129.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., The International Wheat Genome Sequencing Consortium, Jakobsen, K.S., Wulff, B.B.H., Steuernagel, B., Mayer, K.F.X. and Olsen, O.-A. (2014). *Data from: Ancient hybridizations among the ancestral genomes of bread wheat*: Dryad Digital Repository.

Marraccini, P., Rogers, J. W., Allard, C., André, M.-L., Caillet, V., Lacoste, N., Lausanne, F. and Michaux, S. (2001) Molecular and biochemical characterization of endo-β-mannanases from germinating coffee (Coffea arabica) grains. *Planta* **213**, 296–308.

Marulanda, J. J., Mi, X., Melchinger, A. E., Xu, J.-L., Wurschum, T. and Longin, C. F. H. (2016) Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. *TAG Theoretical and Applied Genetics*.

Massawe, F., Mayes, S. and Cheng, A. (2016) Crop Diversity: An Unexploited Treasure Trove for Food Security. *Trends in plant science* **21**, 365–368.

Mayer, K. F. X., Waugh, R., Brown, J. W. S., Schulman, A., Langridge, P., Platzer, M., Fincher, G. B., Muehlbauer, G. J., Sato, K., Close, T. J., Wise, R. P. and Stein, N. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716.

Melchinger, A. E. (1999) Genetic diversity of heterosis in crops. In: *Genetics and exploitation of heterosis in crops* (Gerdes, J.T., ed), pp. 99–118. Madison, WI.

Meuwissen, T. H., Hayes, B. J. and Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.

Mohammadi, S. A. and Prasanna, B. M. (2003) Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Science* **43**, 1235.

Mondal, S., Rutkoski, J. E., Velu, G., Singh, P. K., Crespo-Herrera, L. A., Guzmán, C., Bhavani, S., Lan, C., He, X. and Singh, R. P. (2016) Harnessing Diversity in Wheat to Enhance Grain Yield, Climate Resilience, Disease and Insect Pest Resistance and Nutrition Through Conventional and Modern Breeding Approaches. *Front. Plant Sci.* **7**, 243.

Nelson, D. E., Repetti, P. P., Adams, T. R., Creelman, R. A., Wu, J., Warner, D. C., Anstrom, D. C., Bensen, R. J., Castiglioni, P. P., Donnarummo, M. G., Hinchey, B. S., Kumimoto, R. W., Maszle, D. R., Canales, R. D., Krolikowski, K. A., Dotson, S. B., Gutterson, N., Ratcliffe, O. J. and Heard, J. E. (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance

and lead to improved corn yields on water-limited acres. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16450–16455.

Nielsen, N. H., Backes, G., Stougaard, J., Andersen, S. U. and Jahoor, A. (2014) Genetic diversity and population structure analysis of European hexaploid bread wheat (Triticum aestivum L.) varieties. *PloS one* **9**, e94000.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., Grigoriev, I. V., Lyons, E., Maher, C. A., Martis, M., Narechania, A., Otillar, R. P., Penning, B. W., Salamov, A. A., Wang, Y., Zhang, L., Carpita, N. C., Freeling, M., Gingle, A. R., Hash, C. T., Keller, B., Klein, P., Kresovich, S., McCann, M. C., Ming, R., Peterson, D. G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K. F. X., Messing, J. and Rokhsar, D. S. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551–556.

Payne, P. I., Nightingale, M. A., Krattiger, A. G. and Holt, L. M. (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *J. Sci. Food Agric.* **40**, 51–65.

Poland, J. (2015) Breeding-assisted genomics. *Current opinion in plant biology* **24**, 119–124.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M. and Jannink, J.-L. (2012) Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *The Plant Genome* **5**, 103.

Poland, J. A. and Rife, T. W. (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome* **5**, 92.

Prasad, M., Kumar, N., Kulwal, P. L., Roder, M. S., Balyan, H. S., Dhaliwal, H. S. and Gupta, P. K. (2003) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *TAG Theoretical and Applied Genetics* **106**, 659–667.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, Manuel A R, Bender, D., Maller, J., Sklar, P., de Bakker, Paul I W, Daly, M. J. and Sham, P. C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–575.

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. and Foley, J. A. (2012) Recent patterns of crop yield growth and stagnation. *Nature communications* **3**, 1293.

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., van Ginkel, M., Hoisington, D., Bohn, M. and Melchinger, A. E. (2005) Wheat genetic diversity trends during domestication and breeding. *TAG Theoretical and Applied Genetics* **110**, 859–864.

Ren, Y., He, X., Liu, D., Li, J., Zhao, X., Li, B., Tong, Y., Zhang, A. and Li, Z. (2012) Major quantitative trait loci for seminal root morphology of wheat seedlings. *Mol Breeding* **30**, 139–148.

Roussel, V., Leisova, L., Exbrayat, F., Stehno, Z. and Balfourier, F. (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. *TAG Theoretical and Applied Genetics* **111**, 162–170.

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., Rouse, M. N., Jannink, J.-L. and Sorrells, M. E. (2014) Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *The Plant Genome* **7**, 0.

Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. and Martin, W. (2002) Genetics and geography of wild cereal domestication in the near east. *Nature reviews. Genetics* **3**, 429–441.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S. M., Belter, E., Du, F., Kim, K., Abbott, R. M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S. M., Gillam, B., Chen, W., Le Yan, Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M. J., McMahan, L., van Buren, P., Vaughn, M. W., Ying, K., Yeh, C.-T., Emrich, S. J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W. B., Baucom, R. S., Brutnell, T. P., Carpita, N. C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J. C., Fu, Y., Jeddeloh, J. A., Han, Y., Lee, H., Li, P., Lisch, D. R., Liu, S., Liu, Z., Nagel, D. H., McCann, M. C., SanMiguel, P., Myers, A. M., Nettleton, D., Nguyen, J., Penning, B. W., Ponnala, L., Schneider, K. L., Schwartz, D. C.,

Sharma, A., Soderlund, C., Sprer, N. M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T. K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J. L., Dawe, R. K., Jiang, J., Jiang, N., Presting, G. G., Wessler, S. R., Aluru, S., Martienssen, R. A., Clifton, S. W., McCombie, W. R., Wing, R. A. and Wilson, R. K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science (New York, N.Y.)* **326**, 1112–1115.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., Zhang, X., Rusticucci, M., Semenov, V., Alexander, L. V., Allen, S., Benito, G., Cavazos, T., Clague, J., Conway, D., Della-Marta, P. M., Gerber, M., Gong, S., Goswami, B. N., Hemer, M., Huggel, C., van den Hurk, B., Kharin, V. V., Kitoh, A., Tank, A. M. K., Li, G., Mason, S., McGuire, W., van Oldenborgh, G. J., Orlowsky, B., Smith, S., Thiaw, W., Velegrakis, A., Yiou, P., Zhang, T., Zhou, T. and Zwiers, F. W. (2012) Changes in Climate Extremes and their Impacts on the Natural Physical Environment. In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (Field, C.B., Barros, V., Stocker, T.F. and Dahe, Q., eds), pp. 109–230. Cambridge: Cambridge University Press.

Sharma, S., Xu, S., Ehdaie, B., Hoops, A., Close, T. J., Lukaszewski, A. J. and Waines, J. G. (2011) Dissection of QTL effects for root traits using a chromosome arm-specific mapping population in bread wheat. *TAG Theoretical and Applied Genetics* **122**, 759–769.

Shewry, P. R. (2009) Wheat. *Journal of experimental botany* **60**, 1537–1553.

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M. and Muricho, G. (2013) Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Sec.* **5**, 291–317.

Snowdon, R. J., Abbadi, A., Kox, T., Schmutzer, T. and Leckband, G. (2015) Heterotic Haplotype Capture: precision breeding for hybrid performance. *Trends in plant science* **20**, 410–413.

Spiertz, J. and Ewert, F. (2009) Crop production and resource use to meet the growing demand for food, feed and fuel: Opportunities and constraints. *NJAS - Wageningen Journal of Life Sciences* **56**, 281–300.

Tester, M. and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science (New York, N.Y.)* **327**, 818–822.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796–815.

The International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science (New York, N.Y.)* **345**, 1251788.

Turuspekov, Y., Plieske, J., Ganal, M., Akhunov, E. and Abugalieva, S. (2015) Phylogenetic analysis of wheat cultivars in Kazakhstan based on the wheat 90 K single nucleotide polymorphism array. *Plant Genet. Resour.*, 1–7.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T. M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K. F. and Schon, C.-C. (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC genomics* **15**, 823.

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. and Snowdon, R. J. (2015) Subgenomic Diversity Patterns Caused by Directional Selection in Bread Wheat Gene Pools. *The Plant Genome* **8**, 0.

Voss-Fels, K. and Snowdon, R. J. (2016) Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant biotechnology journal* **14**, 1086–1094.

Waines, J. G. and Ehdaie, B. (2007) Domestication and crop physiology: roots of green-revolution wheat. *Annals of botany* **100**, 991–998.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A. R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M.-C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K. J., Hayden, M. and Akhunov, E. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant biotechnology journal* **12**, 787–796.

Whitford, R., Fleury, D., Reif, J. C., Garcia, M., Okada, T., Korzun, V. and Langridge, P. (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *Journal of experimental botany* **64**, 5411–5428.

Wieckhorst, S., Bekele, W., Kloiber-Maitz, M., Schulz-Streeck, T., Knaak, C., Outunova, M., Davassi, A., Snowdon, R. J. A high-density SNP genotyping array for genome-based breeding of energy sorghum for central Europe. In: *Plant and Animal Genome XXIII 2015, San Diego, CA, USA,* https://pag.confex.com/pag/xxiii/webprogram/Paper15935.html.

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L. A., Benbow, H. R., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., King, J., West, C., Griffiths, S., King, I., Bentley, A. R. and Edwards, K. J. (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant biotechnology journal* **14**, 1195–1206.

Yu, J., Holland, J. B., McMullen, M. D. and Buckler, E. S. (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 539–551.

Zanke, C., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Beier, S., Ganal, M. W. and Roder, M. S. (2014) Genetic architecture of main effect QTL for heading date in European winter wheat. *Frontiers in plant science* **5**, 217.

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, F., Eichhorn, A., Polley, A., Jaenecke, C., Ganal, M. W. and Roder, M. S. (2015) Analysis of main effect QTL for thousand grain weight in European winter wheat (Triticum aestivum L.) by genome-wide association mapping. *Frontiers in plant science* **6**, 644.

Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Wurschum, T., Mock, H.-P., Matros, A., Ebmeyer, E., Schachschneider, R., Kazman, E., Schacht, J., Gowda, M., Longin, C. F. H. and Reif, J. C. (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 15624–15629.

Zhou, W.-C., Kolb, F. L., Bai, G.-H., Domier, L. L., Boze, L. K. and Smith, N. J. (2003) Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breeding* **122**, 40–46.

9 Appendices

Appendix I: Supplementary materials from:

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. and Snowdon, R. J. (2015) Subgenomic Diversity Patterns Caused by Directional Selection in Bread Wheat Gene Pools. *The Plant Genom*e 8, 0. doi: 10.3835/plantgenome2015.03.0013.

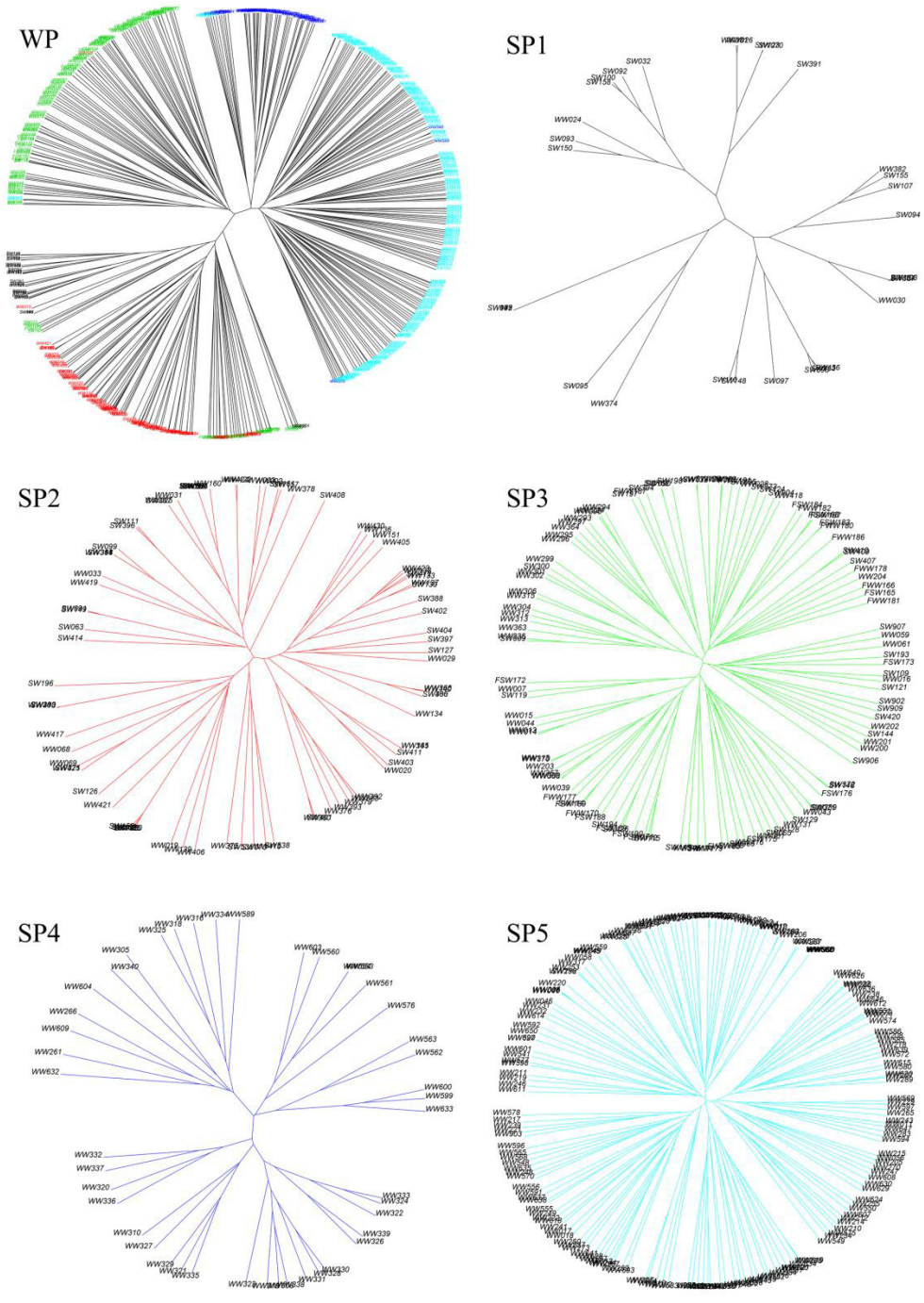


Figure S1. Figure S1. Phylogenic trees for the whole population (WP) and the identified subgroups SP1 to SP5. Unweighted pair group method with arithmetic mean (UPGMA) clustering was performed based on modified Roger's distances polymorphic single-nucleotide-polymorphism (SNP) markers with a minor allele frequency (MAF) ≥ 10%.

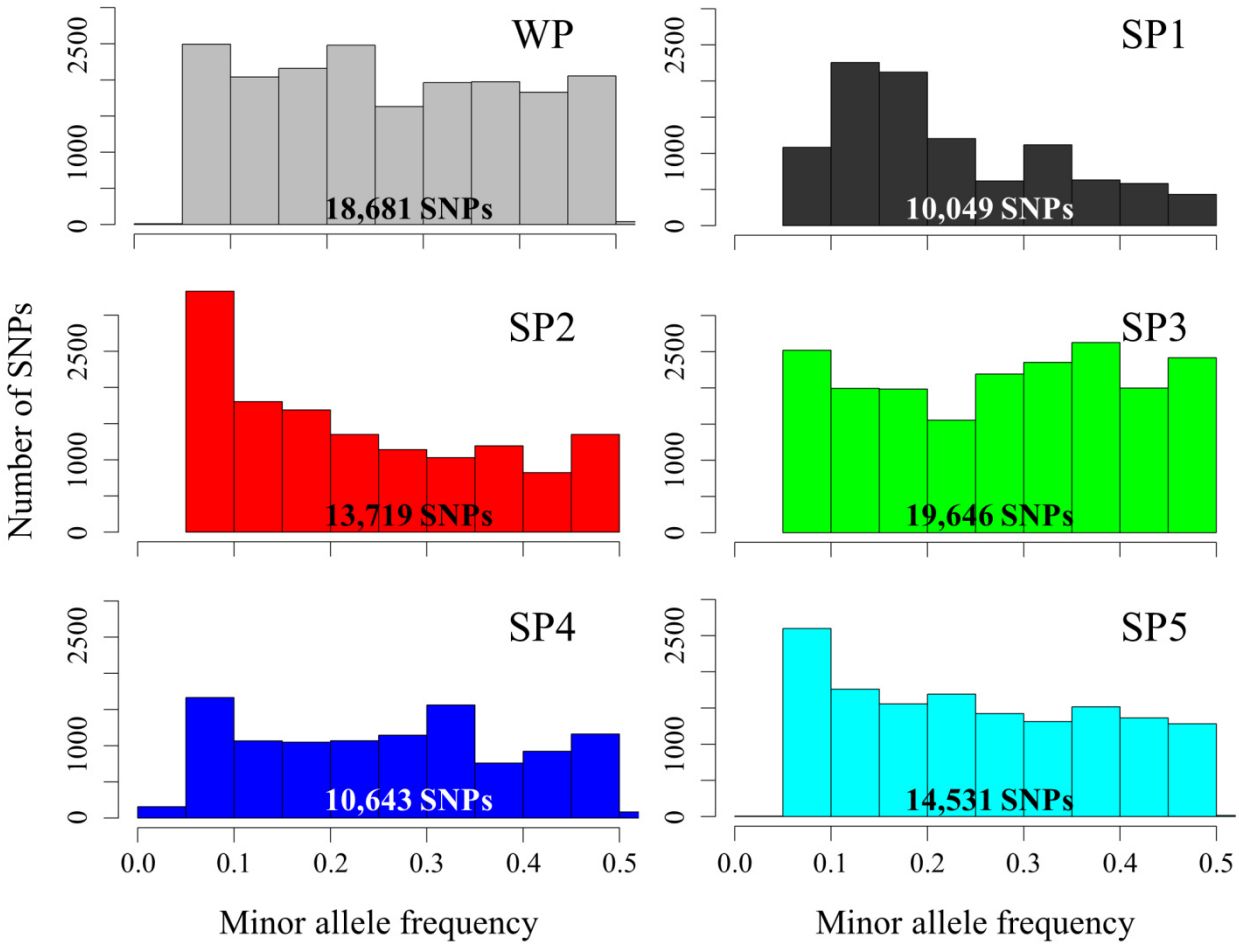


Figure S2. Minor allele frequencies (MAF) of polymorphic single-nucleotide-polymorphism (SNP) markers (MAF ≥ 5%) in the whole population compared to the five subpopulations (SP1-SP5). Numbers in the histograms indicate the number of polymorphic, mapped SNPs in the corresponding germplasm group.

Table S1. Genotype summary of the 460 tested wheat lines

ID-No.	Genotype name	Subgroup assignment based on k-means clustering	Growth type	Time of flowering	Origin	Pedigree (if available)
SW087	WvB/W-SW087	0	spr	early	China	
SW120	WvB/W-SW120	0	spr	early	China	
SW142	Yangmai16	0	spr	early	China	
WW001	WvB/W-WW001	0	win	early	China	
WW023	WvB/W-WW023	0	win	early	China	
WW041	WvB/W-WW041	0	win	early	China	
WW195	WvB/W-WW195	0	win	early	China	
WW566	WvB/W-WW566	0	win	late	Germany	
WW571	WvB/W-WW571	0	win	late	Germany	
WW573	WvB/W-WW573	0	win	late	Germany	
SW032	WvB/W-SW032	1	spr	early	China	
SW070	WvB/W-SW070	1	spr	early	China	
SW072	WvB/W-SW072	1	spr	early	China	
SW092	Ningmai13	1	spr	early	China	
SW093	Ningmai14	1	spr	early	China	
SW094	Ning0310	1	spr	early	China	
SW095	Ning03119	1	spr	early	China	
SW097	WvB/W-SW097	1	spr	early	China	
SW098	WvB/W-SW098	1	spr	early	China	
SW100	WvB/W-SW100	1	spr	early	China	
SW102	Wangshubai	1	spr	early	China	
SW107	WvB/W-SW107	1	spr	early	China	
SW110	WvB/W-SW110	1	spr	early	China	
SW123	WvB/W-SW123	1	spr	early	China	
SW147	Wangshubai	1	spr	early	China	
SW148	WvB/W-SW148	1	spr	early	China	
SW150	WvB/W-SW150	1	spr	early	China	
SW153	WvB/W-SW153	1	spr	early	China	
SW155	WvB/W-SW155	1	spr	early	China	
SW156	WvB/W-SW156	1	spr	early	China	
SW158	WvB/W-SW158	1	spr	early	China	
SW164	WvB/W-SW164	1	spr	early	China	
SW391	WvB/W-SW391	1	spr	early	China	
SW537	WvB/W-SW537	1	spr	early	China	
SW905	Wangshubai	1	spr	early	China	
WW024	WvB/W-WW024	1	win	early	China	
WW026	WvB/W-WW026	1	win	early	China	
WW030	WvB/W-WW030	1	win	early	China	
WW103	Shengxuan4	1	win	early	China	
WW374	Ning05562	1	win	early	China	
WW381	Ning0569	1	win	early	China	
WW382	WvB/W-WW382	1	win	early	China	
SW063	WvB/W-SW063	2	spr	early	China	
SW096	WvB/W-SW096	2	spr	early	China	
SW099	WvB/W-SW099	2	spr	early	China	
SW111	WvB/W-SW111	2	spr	early	China	
SW125	WvB/W-SW125	2	spr	early	China	
SW126	WvB/W-SW126	2	spr	early	China	
SW127	WvB/W-SW127	2	spr	early	China	
SW130	WvB/W-SW130	2	spr	early	China	
SW149	WvB/W-SW149	2	spr	early	China	

SW152	WvB/W-SW152	2	spr	early	China	
SW157	WvB/W-SW157	2	spr	early	China	
SW159	WvB/W-SW159	2	spr	early	China	
SW161	WvB/W-SW161	2	spr	early	China	
SW163	WvB/W-SW163	2	spr	early	China	
SW371	Xumai29	2	spr	early	China	
SW384	WvB/W-SW384	2	spr	early	China	
SW385	WvB/W-SW385	2	spr	early	China	
SW386	WvB/W-SW386	2	spr	early	China	
SW387	Yang 06G5	2	spr	early	China	
SW388	WvB/W-SW388	2	spr	early	China	
SW390	Zhe0616	2	spr	early	China	
SW396	WvB/W-SW396	2	spr	early	China	
SW397	WvB/W-SW397	2	spr	early	China	
SW398	WvB/W-SW398	2	spr	early	China	
SW400	WvB/W-SW400	2	spr	early	China	
SW401	WvB/W-SW401	2	spr	early	China	
SW402	WvB/W-SW402	2	spr	early	China	
SW403	Ningnuo 2	2	spr	early	China	
SW404	WvB/W-SW404	2	spr	early	China	
SW408	WvB/W-SW408	2	spr	early	China	
SW411	Shibin14	2	spr	early	China	
SW412	WvB/W-SW412	2	spr	early	China	
SW413	Shibin5	2	spr	early	China	
SW414	WvB/W-SW414	2	spr	early	China	
SW536	WvB/W-SW536	2	spr	early	China	
SW538	WvB/W-SW538	2	spr	early	China	
SW73	WvB/W-SW73	2	spr	early	China	
SW91	Ningmai 12	2	spr	early	China	
WW002	WvB/W-WW002	2	win	early	China	
WW003	WvB/W-WW003	2	win	early	China	
WW020	WvB/W-WW020	2	win	early	China	
WW025	WvB/W-WW025	2	win	early	China	
WW029	WvB/W-WW029	2	win	early	China	
WW031	WvB/W-WW031	2	win	early	China	
WW033	WvB/W-WW033	2	win	early	China	
WW068	WvB/W-WW068	2	win	early	China	
WW069	WvB/W-WW069	2	win	early	China	
WW133	Yangmai158	2	win	early	China	
WW134	Yangmai17	2	win	early	China	
WW135	Ningyan1	2	win	early	China	
WW136	Ningnuo1	2	win	early	China	
WW137	Yangmai11	2	win	early	China	
WW138	WvB/W-WW138	2	win	early	China	
WW139	Yangmai13	2	win	early	China	
WW140	Yangmai14	2	win	early	China	
WW141	Yangmai15	2	win	early	China	
WW151	WvB/W-WW151	2	win	early	China	
WW154	WvB/W-WW154	2	win	early	China	
WW160	WvB/W-WW160	2	win	early	China	
WW19	WvB/W-WW19	2	win	early	China	
WW196	Chuanyu12	2	win	early	China	
WW303	W-WW303	2	win	late	Netherlands	TABASCO x SHA3/CBRD
WW311	W-WW311	2	win	late	Netherlands	TABASCO x SHA3/CBRD
WW360	WvB/W-WW360	2	win	early	China	
WW365	WvB/W-WW365	2	win	early	China	
WW366	WvB/W-WW366	2	win	early	China	

WW375	Ning0762	2	win	early	China
WW376	Ning0644	2	win	early	China
WW378	Ning0604	2	win	early	China
WW379	Ning0588	2	win	early	China
WW380	Ning0575	2	win	early	China
WW383	Ningmai14	2	win	early	China
WW392	WvB/W-WW392	2	win	early	China
WW393	WvB/W-WW393	2	win	early	China
WW405	WvB/W-WW405	2	win	early	China
WW406	WvB/W-WW406	2	win	early	China
WW415	Sumai 2	2	win	early	China
WW417	WvB/W-WW417	2	win	early	China
WW419	WvB/W-WW419	2	win	early	China
WW421	WvB/W-WW421	2	win	early	China
WW422	WvB/W-WW422	2	win	early	China
WW423	WvB/W-WW423	2	win	early	China
WW424	WvB/W-WW424	2	win	early	China
WW425	WvB/W-WW425	2	win	early	China
WW430	WvB/W-WW430	2	win	early	China
FSW165	Shijiazhuang8	3	fac spr	early	China
FSW167	Zhoumai16	3	fac spr	early	China
FSW168	Zhoumai18	3	fac spr	early	China
FSW169	Jimai19	3	fac spr	early	China
FSW171	Jimai22	3	fac spr	early	China
FSW172	Zhengmai366	3	fac spr	early	China
FSW173	Hebeinongda341	3	fac spr	early	China
FSW174	Ji3475	3	fac spr	early	China
FSW175	Kaocheng8901	3	fac spr	early	China
FSW176	Gaoyou503	3	fac spr	early	China
FSW183	Zhou91177	3	fac spr	early	China
FSW184	Zhoumai13	3	fac spr	early	China
FSW188	Shannong990525	3	fac spr	early	China
FSW189	Yan2801	3	fac spr	early	China
FSW190	Yan475	3	fac spr	early	China
FWW166	Han6172	3	fac win	early	China
FWW177	Zhongyu6	3	fac win	early	China
FWW178	Zhong6	3	fac win	early	China
FWW179	98 Zhong 18	3	fac win	early	China
FWW180	Yumai35	3	fac win	early	China
FWW181	Yumai50	3	fac win	early	China
FWW182	Yumai62	3	fac win	early	China
FWW185	WvB/W-FWW185	3	fac win	early	China
FWW186	Lankao24	3	fac win	early	China
FWW187	WvB/W-FWW187	3	fac win	early	China
FWW191	Shan302518	3	fac win	early	China
SW065	WvB/W-SW065	3	spr	early	China
SW066	WvB/W-SW066	3	spr	early	China
SW071	WvB/W-SW071	3	spr	early	China
SW074	Sumai1	3	spr	early	China
SW090	Ningmai 11	3	spr	early	China
SW101	WvB/W-SW101	3	spr	early	China
SW108	WvB/W-SW108	3	spr	early	China
SW109	WvB/W-SW109	3	spr	early	China
SW112	WvB/W-SW112	3	spr	early	China
SW114	WvB/W-SW114	3	spr	early	China
SW115	W-SW115	3	spr	early	China
SW119	WvB/W-SW119	3	spr	early	China

SW121	WvB/W-SW121	3	spr	early	China	
SW124	WvB/W-SW124	3	spr	early	China	
SW128	WvB/W-SW128	3	spr	early	China	
SW129	WvB/W-SW129	3	spr	early	China	
SW144	Sumai3-WvB	3	spr	early	China	
SW145	Xuzhou 1	3	spr	early	China	
SW146	Xuzhou 2	3	spr	early	China	
SW162	WvB/W-SW162	3	spr	early	China	
SW193	Huaimai18	3	spr	early	China	
SW194	Wanmai38	3	spr	early	China	
SW197	WvB/W-SW197	3	spr	early	China	
SW198	Yunmai44	3	spr	early	China	
SW300	W-SW300	3	spr	late	Netherlands	W x SHA3/CBRD
SW309	W-SW309	3	spr	late	Netherlands	TABASCO x SHA3/CBRD
SW372	Xumai27	3	spr	early	China	
SW373	Ning07233	3	spr	early	China	
SW389	Mai 48	3	spr	early	China	
SW394	WvB/W-SW394	3	spr	early	China	
SW395	WvB/W-SW395	3	spr	early	China	
SW407	WvB/W-SW407	3	spr	early	China	
SW409	WvB/W-SW409	3	spr	early	China	
SW410	WvB/W-SW410	3	spr	early	China	
SW420	WvB/W-SW420	3	spr	early	China	
SW426	J95	3	spr	early	China	
SW902	Florence Aurore	3	spr	early	France	
SW904	WvB/W-SW904	3	spr	early	China	
SW906	CSCR6	3	spr	early	Australia	
SW907	CSCR14	3	spr	early	Australia	
SW908	CSCR16	3	spr	early	Australia	
SW909	CSCR28	3	spr	early	Australia	
WW/F170	Jimai20	3	fac win	early	China	
WW007	WvB/W-WW007	3	win	early	China	
WW028	WvB/W-WW028	3	win	early	China	
WW038	WvB/W-WW038	3	win	early	China	
WW039	WvB/W-WW039	3	win	early	China	
WW043	WvB/W-WW043	3	win	early	China	
WW044	WvB/W-WW044	3	win	early	China	
WW059	WvB/W-WW059	3	win	early	North America	
WW061	WvB/W-WW061	3	win	late	Slovakia	
WW113	WvB/W-WW113	3	win	early	China	
WW13	WvB/W-WW13	3	win	early	China	
WW131	WvB/W-WW131	3	win	early	China	
WW132	WvB/W-WW132	3	win	early	China	
WW14	WvB/W-WW14	3	win	early	China	
WW15	WvB/W-WW15	3	win	early	China	
WW16	WvB/W-WW16	3	win	early	China	
WW200	WvB/W-WW200	3	win	early	China	
WW201	WvB/W-WW201	3	win	early	China	
WW202	Jingdong17	3	win	early	China	
WW203	Jinmai61	3	win	early	China	
WW204	Jinong207	3	win	early	China	
WW224	Intro	3	win	late	Germany	
WW235	Linus	3	win	late	Germany	
WW293	W-WW293	3	win	late	Netherlands	W x NG8675/CBRD
WW294	W-WW294	3	win	late	Netherlands	W x NG8675/CBRD
WW295	W-WW295	3	win	late	Netherlands	W x NG8675/CBRD
WW296	W-WW296	3	win	late	Netherlands	W x NG8675/CBRD

WW297	W-WW297	3	win	late	Netherlands	W x NG8675/CBRD
WW298	W-WW298	3	win	late	Netherlands	W x NG8675/CBRD
WW299	W-WW299	3	win	late	Netherlands	W x SHA3/CBRD
WW301	W-WW301	3	win	late	Germany	W x SHA3/CBRD
WW302	W-WW302	3	win	late	Netherlands	W x SHA3/CBRD
WW304	W-WW304	3	win	late	Netherlands	TABASCO x SHA3/CBRD
WW306	W-WW306	3	win	late	Netherlands	TABASCO x SHA3/CBRD
WW312	W-WW312	3	win	late	Netherlands	TABASCO x SHA3/CBRD
WW313	W-WW313	3	win	late	Netherlands	TABASCO x SHA3/CBRD
WW315	W-WW315	3	win	late	Netherlands	TABASCO x SHA3/CBRD
WW361	WvB/W-WW361	3	win	early	China	
WW363	WvB/W-WW363	3	win	early	China	
WW364	WvB/W-WW364	3	win	early	China	
WW367	WvB/W-WW367	3	win	early	China	
WW368	Huaimai23	3	win	early	China	
WW369	Huaimai 21	3	win	early	China	
WW370	Huaimai 20	3	win	early	China	
WW416	WvB/W-WW416	3	win	early	China	
WW418	WvB/W-WW418	3	win	early	China	
WW545	JLU-WW545	3	win	late	Germany	
WW261	Anapolis	4	win	late	Germany	
WW266	Forum	4	win	late	Germany	
WW305	W-WW305	4	win	late	Netherlands	TABASCO x SHA3/CBRD
WW308	W-WW308	4	win	late	Netherlands	TABASCO x SHA3/CBRD
WW310	W-WW310	4	win	late	Netherlands	TABASCO x SHA3/CBRD
WW316	W-WW316	4	win	late	Netherlands	ANTHUS x PAMIER
WW318	W-WW318	4	win	late	Netherlands	W x SHA3/CBRD
WW319	W-WW319	4	win	late	Netherlands	W x SHA3/CBRD
WW320	W-WW320	4	win	late	Netherlands	W x SHA3/CBRD
WW321	W-WW321	4	win	late	Netherlands	W x SHA3/CBRD
WW322	W-WW322	4	win	late	Netherlands	W x SHA3/CBRD
WW323	W-WW323	4	win	late	Netherlands	W x SHA3/CBRD
WW324	W-WW324	4	win	late	Netherlands	W x SHA3/CBRD
WW325	W-WW325	4	win	late	Netherlands	W x SHA3/CBRD
WW326	W-WW326	4	win	late	Netherlands	W x SHA3/CBRD
WW327	W-WW327	4	win	late	Netherlands	W x SHA3/CBRD
WW328	W-WW328	4	win	late	Netherlands	W x SHA3/CBRD
WW329	W-WW329	4	win	late	Netherlands	W x SHA3/CBRD
WW330	W-WW330	4	win	late	Netherlands	W x SHA3/CBRD
WW331	W-WW331	4	win	late	Netherlands	W x SHA3/CBRD
WW332	W-WW332	4	win	late	Netherlands	W x SHA3/CBRD
WW333	W-WW333	4	win	late	Netherlands	W x SHA3/CBRD
WW334	W-WW334	4	win	late	Netherlands	W x SHA3/CBRD
WW335	W-WW335	4	win	late	Netherlands	W x SHA3/CBRD
WW336	W-WW336	4	win	late	Netherlands	W x SHA3/CBRD
WW337	W-WW337	4	win	late	Netherlands	W x SHA3/CBRD
WW338	W-WW338	4	win	late	Netherlands	W x SHA3/CBRD
WW339	W-WW339	4	win	late	Netherlands	W x SHA3/CBRD
WW340	W-WW340	4	win	late	Netherlands	W x SHA3/CBRD
WW553	WvB/W-WW553	4	win	late	Germany	
WW554	WvB/W-WW554	4	win	late	Germany	
WW560	WvB/W-WW560	4	win	late	Germany	
WW561	WvB/W-WW561	4	win	late	Germany	
WW562	WvB/W-WW562	4	win	late	Germany	
WW563	WvB/W-WW563	4	win	late	Germany	
WW576	WvB/W-WW576	4	win	late	Germany	
WW589	WvB/W-WW589	4	win	late	Germany	

WW599	WvB/W-WW599	4	win	late	Germany	
WW600	WvB/W-WW600	4	win	late	Germany	
WW603	WvB/W-WW603	4	win	late	Germany	
WW604	WvB/W-WW604	4	win	late	Germany	
WW609	WvB/W-WW609	4	win	late	Germany	
WW632	WvB/W-WW632	4	win	late	Germany	
WW633	WvB/W-WW633	4	win	late	Germany	
WW004	WvB/W-WW004	5	win	early	China	
WW005	WvB/W-WW005	5	win	early	China	
WW006	WvB/W-WW006	5	win	early	China	
WW010	WvB/W-WW010	5	win	early	China	
WW011	Tobak	5	win	late	Germany	(Ellvis x Drifter) x Koch
WW017	WvB/W-WW017	5	win	early	China	
WW046	Apache	5	win	late	France	
WW047	WvB-WW047	5	win	late	France	
WW049	Premio	5	win	late	France	
WW056	WvB-WW056	5	win	late	Netherlands	
WW058	WvB-WW058	5	win	late	Netherlands	
WW060	WvB-WW060	5	win	late	Romania	
WW12	WvB/W-WW12	5	win	early	China	
WW18	WvB/W-WW18	5	win	early	China	
WW192	Huaimai17	5	win	early	China	
WW199	Yunmai46	5	win	early	China	
WW205	WvB-WW205	5	win	late	Germany	(Ilias x Darwin) x Koch
WW206	WvB-WW206	5	win	late	Germany	Tyberius x Opus
WW207	WvB-WW207	5	win	late	Germany	(Qualibo x Tommi) x Tulsa
WW208	WvB-WW208	5	win	late	Germany	(China 1 x Opus) x Tulsa
WW209	Akteur	5	win	late	Germany	
WW210	Alchemy	5	win	late	France	
WW211	Alves	5	win	late	UK	
WW212	Beluga	5	win	late	UK	
WW213	Cassius	5	win	late	UK	
WW214	Claire	5	win	late	UK	
WW215	Conqueror	5	win	late	UK	
WW217	Egoist	5	win	late	Germany	(10165 H x EC 46-26) x 10373-9
WW218	Elixer	5	win	late	Germany	(Semper x Bristol) x Tulsa
WW219	Esket	5	win	late	Germany	
WW220	Genius	5	win	late	Germany	
WW221	Gladiator	5	win	late	UK	
WW222	Global	5	win	late	Germany	
WW223	Inspiration	5	win	late	Germany	
WW225	JB Asano	5	win	late	Germany	
WW226	Julius	5	win	late	Germany	
WW227	Kerubino	5	win	late	Germany	
WW228	Kometus	5	win	late	Germany	
WW229	Kredo	5	win	late	Germany	
WW230	KWS Dacanto	5	win	late	Germany	
WW231	KWS Ferrum	5	win	late	Germany	
WW232	KWS Sterling	5	win	late	UK	
WW234	Lear	5	win	late	Germany	
WW236	Lucius	5	win	late	Germany	
WW237	Meister	5	win	late	Germany	
WW238	Muskat	5	win	late	Germany	ZE 21372 x 86Z99-9
WW239	Nelson	5	win	late	Germany	
WW241	Orcas	5	win	late	Germany	
WW242	Panorama	5	win	late	UK	
WW243	Pengar	5	win	late	Germany	(Ellvis x Drifter) x Levendis

WW244	Potenzial	5	win	late	Germany	
WW245	Premio	5	win	late	Germany	
WW246	Sailor	5	win	late	Germany	
WW247	Scout	5	win	late	UK	
WW248	Skalmeje	5	win	late	Germany	
WW249	Sokrates	5	win	late	Germany	
WW250	Stigg	5	win	late	UK	
WW253	Tabasco	5	win	late	Germany	(ZE.90-2666 x 86Z99.9) x CPB.93-27
WW255	Warrior	5	win	late	UK	
WW256	Zappa	5	win	late	Germany	(ZE.90-2666 x 86Z99.9) x CPB.93-27
WW258	JLU-WW258	5	win	late	China	
WW263	Mentor	5	win	late	Germany	
WW265	SY Ferry	5	win	late	Germany	
WW267	Pionier	5	win	late	Germany	
WW268	Apian	5	win	late	Germany	
WW269	Gordian	5	win	late	Germany	
WW270	Desamo	5	win	late	Germany	
WW271	Edward	5	win	late	Germany	(Corvus x Bristol) x Biscay
WW272	Boxer	5	win	late	Germany	Idol x Tommi
WW273	Memory	5	win	late	Germany	
WW283	WvB-WW283	5	win	late	Germany	Idol x Tommi
WW286	WB 564209	5	win	late	Germany	(Robigus x Sobi) x Tulsa
WW287	WvB-WW287	5	win	late	Germany	(Roswell x Privileg) x Bristol
WW288	BB 619609	5	win	late	Germany	Hermann x Tulsa
WW289	BB 690909	5	win	late	Germany	(Qualibo x Tommi) x Tulsa
WW290	BB-732009-W	5	win	late	Germany	Türkis x Hermann
WW292	W-WW292	5	win	late	Netherlands	W x NG8675/CBRD
WW307	W-WW307	5	win	late	Netherlands	TABASCO x SHA3/CBRD
WW314	W-WW314	5	win	late	Netherlands	TABASCO x SHA3/CBRD
WW317	W-WW317	5	win	late	Netherlands	W x SHA3/CBRD
WW532	WvB/JLU-WW532	5	win	late	China	
WW540	Colonia	5	win	late	Germany	
WW541	Patras	5	win	late	Germany	
WW542	KWS Montana	5	win	late	Germany	
WW543	Rumor	5	win	late	Germany	
WW547	WvB/JLU-WW547	5	win	late	Germany	
WW548	WvB/W-WW548	5	win	late	Germany	
WW549	Accroc	5	win	late	France	
WW550	Alixan	5	win	late	France	
WW551	Sachsmo	5	win	late	Germany	
WW552	WvB/W-WW552	5	win	late	Germany	
WW555	WvB/W-WW555	5	win	late	Germany	
WW556	WvB/W-WW556	5	win	late	Germany	
WW557	WvB/W-WW557	5	win	late	Germany	
WW558	WvB/W-WW558	5	win	late	Germany	
WW559	WvB/W-WW559	5	win	late	Germany	
WW564	WvB/W-WW564	5	win	late	Germany	
WW565	WvB/W-WW565	5	win	late	Germany	
WW567	WvB/W-WW567	5	win	late	Germany	
WW568	WvB/W-WW568	5	win	late	Germany	
WW569	WvB/W-WW569	5	win	late	Germany	
WW570	WvB/W-WW570	5	win	late	Germany	
WW572	WvB/W-WW572	5	win	late	Netherlands	
WW574	WvB/W-WW574	5	win	late	Germany	
WW575	WvB/W-WW575	5	win	late	Germany	
WW577	WvB/W-WW577	5	win	late	Germany	

WW578	WvB/W-WW578	5	win	late	Germany
WW579	WvB/W-WW579	5	win	late	Germany
WW580	WvB/W-WW580	5	win	late	Germany
WW581	WvB/W-WW581	5	win	late	Germany
WW582	WvB/W-WW582	5	win	late	Germany
WW583	WvB/W-WW583	5	win	late	Germany
WW584	WvB/W-WW584	5	win	late	Germany
WW585	WvB/W-WW585	5	win	late	Germany
WW586	WvB/W-WW586	5	win	late	Germany
WW587	WvB/W-WW587	5	win	late	Germany
WW588	WvB/W-WW588	5	win	late	Germany
WW590	WvB/W-WW590	5	win	late	Germany
WW591	WvB/W-WW591	5	win	late	Germany
WW592	WvB/W-WW592	5	win	late	Germany
WW593	WvB/W-WW593	5	win	late	Germany
WW594	WvB/W-WW594	5	win	late	Germany
WW595	WvB/W-WW595	5	win	late	Germany
WW596	WvB/W-WW596	5	win	late	Germany
WW597	WvB/W-WW597	5	win	late	Germany
WW598	WvB-WW598	5	win	late	Germany
WW601	WvB/W-WW601	5	win	late	Germany
WW602	WvB/W-WW602	5	win	late	Germany
WW605	WvB/W-WW605	5	win	late	Germany
WW606	WvB/W-WW606	5	win	late	Germany
WW607	WvB/W-WW607	5	win	late	Germany
WW608	WvB/W-WW608	5	win	late	Germany
WW610	WvB/W-WW610	5	win	late	Germany
WW611	WvB/W-WW611	5	win	late	Germany
WW612	WvB/W-WW612	5	win	late	Germany
WW613	WvB/W-WW613	5	win	late	Germany
WW614	WvB/W-WW614	5	win	late	Germany
WW615	WvB/W-WW615	5	win	late	Germany
WW616	WvB/W-WW616	5	win	late	Germany
WW617	WvB/W-WW617	5	win	late	Germany
WW618	WvB/W-WW618	5	win	late	Germany
WW619	WvB/W-WW619	5	win	late	Germany
WW620	WvB/W-WW620	5	win	late	Germany
WW621	WvB/W-WW621	5	win	late	Germany
WW622	WvB/W-WW622	5	win	late	Germany
WW623	WvB/W-WW623	5	win	late	Germany
WW624	WvB/W-WW624	5	win	late	Germany
WW625	WvB/W-WW625	5	win	late	Germany
WW626	WvB/W-WW626	5	win	late	Germany
WW627	WvB/W-WW627	5	win	late	Germany
WW628	WvB/W-WW628	5	win	late	Germany
WW629	WvB/W-WW629	5	win	late	Germany
WW630	WvB/W-WW630	5	win	late	Germany
WW631	KWS Milaneco	5	win	late	Germany
WW634	WvB/W-WW634	5	win	late	Germany
WW635	WvB/W-WW635	5	win	late	Germany
WW636	WvB/W-WW636	5	win	late	Germany
WW637	WvB/W-WW637	5	win	late	Germany
WW638	WvB/W-WW638	5	win	late	Germany
WW639	WvB/W-WW639	5	win	late	Germany
WW640	WvB/W-WW640	5	win	late	Germany
WW641	WvB/W-WW641	5	win	late	Germany
WW642	WvB/W-WW642	5	win	late	Germany

WW643	WvB/W-WW643	5	win	late	Germany
WW644	WvB/W-WW644	5	win	late	Germany
WW645	WvB/W-WW645	5	win	late	Germany
WW646	WvB/W-WW646	5	win	late	Germany
WW647	WvB/W-WW647	5	win	late	Germany
WW648	WvB/W-WW648	5	win	late	Germany
WW649	WvB-WW649	5	win	late	Germany
WW650	WvB/W-WW650	5	win	late	Germany
WW900	WvB-WW900	5	win	late	Germany
WW901	Dream	5	win	late	Germany
WW903	Lynx	5	win	late	UK

Genotype name: WvB = Breeding line from W. von Borries-Eckendorf GmbH & Co. KG;
W = Breeding line from Wiersum Plant Breeding; JLU = Breeding line from Justus Liebig University

Subgroup assignement: 0 = Genotype exluded from the data set due to filtering restrictions (see Materials and Methods).

Table S2. Number and sizes of gaps between polymorphic SNP markers (MAF ≥ 5%) in the A, B and D subgenome.

Gap size (cM)	Subgenome		
	A	B	D
0	3.473	7.979	1.388
0 - 0.1	137	223	40
0.1 - 0.5	502	672	109
0.5 - 1	310	320	90
1 - 2	209	181	90
2 - 5	117	90	76
5 - 10	18	20	45
10 - 15	-	2	13
15 - 20	2	1	10
20 - 30	1	-	4
> 30	-	-	2
Biggest gap:	21.22	15.77	34.59

Table S3. Average expected heterozigocity (ExHet) and linkage disequilibrium (r2) in the A, B and D subgenome for the whole population (WP) and the five subpopulations (SP) calculated with polymorphic SNP markers (MAF ≥ 5%).

Chromosome		WP		SP1		SP2		SP3		SP4		SP5	
		ExHet	**r^2**	**ExHet**	**r^2**	**ExHet**	**r^2**	**ExHet**	**r^2**	**ExHet**	**r^2**	**ExHet**	**r^2**
A Genome	1A	0.53	0.046	0.25	0.517	0.3	0.129	0.59	0.062	0.32	0.116	0.48	0.056
	2A	0.57	0.071	0.28	0.367	0.46	0.111	0.59	0.078	0.33	0.12	0.48	0.048
	3A	0.55	0.045	0.29	0.248	0.47	0.091	0.58	0.062	0.38	0.17	0.42	0.053
	4A	0.54	0.062	0.34	0.256	0.46	0.103	0.58	0.064	0.3	0.242	0.4	0.088
	5A	0.57	0.046	0.34	0.182	0.41	0.104	0.6	0.06	0.49	0.17	0.48	0.056
	6A	0.59	0.086	0.25	0.295	0.29	0.162	0.58	0.065	0.35	0.148	0.45	0.052
	7A	0.6	0.036	0.38	0.234	0.43	0.122	0.6	0.042	0.35	0.12	0.57	0.046
	Mean	**0.56**	**0.056**	**0.3**	**0.295**	**0.4**	**0.113**	**0.59**	**0.062**	**0.36**	**0.146**	**0.47**	**0.053**
B Genome	1B	0.54	0.129	0.22	0.262	0.3	0.103	0.57	0.143	0.35	0.371	0.36	0.057
	2B	0.61	0.049	0.34	0.24	0.45	0.081	0.6	0.054	0.39	0.152	0.5	0.05
	3B	0.58	0.038	0.31	0.21	0.38	0.078	0.61	0.048	0.29	0.224	0.48	0.052
	4B	0.58	0.068	0.38	0.313	0.44	0.178	0.57	0.061	0.34	0.256	0.47	0.09
	5B	0.56	0.059	0.37	0.302	0.45	0.084	0.57	0.06	0.39	0.148	0.46	0.088
	6B	0.57	0.071	0.24	0.473	0.38	0.114	0.59	0.076	0.3	0.288	0.47	0.085
	7B	0.6	0.052	0.3	0.346	0.42	0.088	0.6	0.049	0.36	0.123	0.46	0.044
	Mean	**0.58**	**0.074**	**0.31**	**0.296**	**0.4**	**0.093**	**0.59**	**0.078**	**0.35**	**0.223**	**0.46**	**0.067**
D Genome	1D	0.44	0.161	0.22	0.356	0.31	0.161	0.47	0.156	0.32	0.25	0.37	0.194
	2D	0.54	0.188	0.19	0.259	0.43	0.277	0.57	0.192	0.24	0.176	0.3	0.078
	3D	0.49	0.107	0.2	0.298	0.3	0.236	0.47	0.107	0.47	0.263	0.46	0.169
	4D	0.57	0.119	0.21	0.841	0.41	0.26	0.58	0.094	0.26	0.217	0.51	0.146
	5D	0.58	0.072	0.33	0.226	0.43	0.107	0.56	0.076	0.34	0.205	0.52	0.079
	6D	0.6	0.152	0.24	0.313	0.38	0.136	0.57	0.119	0.31	0.225	0.5	0.125
	7D	0.48	0.056	0.24	0.185	0.32	0.084	0.47	0.052	0.28	0.181	0.41	0.057
	Mean	**0.53**	**0.163**	**0.23**	**0.306**	**0.37**	**0.241**	**0.53**	**0.161**	**0.32**	**0.226**	**0.44**	**0.147**
	Whole genome	**0.56**	**0.071**	**0.28**	**0.296**	**0.39**	**0.105**	**0.57**	**0.075**	**0.34**	**0.195**	**0.46**	**0.064**

Table S4. Mean r^2 values within candidate regions for directional selection.

Chromosome	Position	WP	SP1	SP2	SP3	SP4	SP5
1A	154.06	0.247	0.534	0.422	0.204	0.243	0.297
2A	25.97	0.289	0.367	0.208	0.19	0.41	0.347
2A	47.22	0.435	1†	0.593	0.427	0.425	0.448
5A	35034	0.072	0.015	0.127	0.0002	0.036	0.003
5A	140.59	0.338	0.68	0.401	0.2	0.845	1
6A	60.27	0.229	1†	1†	0.221	0.214	0.416
6A	79.08	0.459	0.97	0.875	0.413	0.666	0.798
6A	84.66-85.01	0.248	0.543	0.946	0.333	0.671	0.359
1B	60.62	0.786	0.483	0.395	0.827	0.982	0.215
1B	64.89-68.04	0.352	0.608	0.467	0.293	0.81	0.457
1B	70.08	0.2	0.855	0.465	0.241	0.697	0.253
4B	64.58	0.491	0.879	0.652	0.381	0.668	0.514
4B	109.51-110.84	0.104	1	0.297	0.173	1†	0.259
6B	115.25-122.92	0.279	0.278	0.622	0.23	0.426	0.409
7B	10.06	0.727	1†	1†	0.362	0.582	0.431
7B	55.64	0.405	1	0.304	0.32	0.959	0.309
7B	72.74-72.99	0.302	0.637	0.764	0.287	0.938	0.613
1D	39.48	1	1†	1	1	0.731	0.179
1D	45.44	0.141	0.649	0.336	0.143	0.338	0.257
1D	82.05	0.27	0.341	0.403	0.093	1†	0.363
5D	137.88	0.427	0.469	0.273	0.426	1†	0.348
6D	133.61	0.786	1	0.467	0.595	1	0.983
7D	26.33	0.502	1†	1	0.491	0.945	0.591
7D	149.59-152.29	0.421	0.062	0.583	0.373	1†	0.266

r^2 values were calculated in 2 cM windows around single FST-outlier loci. For regions harbouring more than one marker, the mean LD was calculated for the whole section plus 1 additional cM on each flanking side.

†: No polymorphic markers mapped in these regions; r^2 was therefore assumed to be 1.

Table S5. Summary of 150 candidate loci for directional selection found by an Fst outlier detection method.

SNP ID	Alleles	Chromosome	Position (cM)	Strand	Fst-Value	ExHet	Distribution of marker allelles (i/ii) in the whole population (WP) and the subpopulations (SP)						
							SP1	SP2	SP3	SP4	SP5	WP	Total marker hits
wsnp_RFL_Contig3881_4265086	A/G	1A	154.06	U	0.167155	0.067836	0/32	0/85	0/115	0/44	29/142	29/418	447
TA001766-2030	A/G	2A	25.97	U	0.833346	0.566396	32/0	83/1	104/9	0/44	32/142	251/196	447
BS00039973_51	C/T	2A	25.97	S	0.819926	0.561588	32/0	83/1	103/9	0/42	32/120	250/172	422
RAC875_s118883_99	A/G	2A	47.22	S	0.836753	0.567414	32/0	84/1	105/9	0/44	31/143	252/197	449
BS00086365_51	A/G	2A	47.22	S	0.834605	0.568318	0/32	1/83	9/96	44/0	142/30	196/241	437
BS00014736_51	C/T	2A	47.22	U	0.83426	0.566346	32/0	84/1	106/9	0/44	32/142	254/196	450
Kukri_c24446_306	A/C	2A	47.22	S	0.833758	0.566372	32/0	83/1	105/9	0/44	32/142	252/196	448
Kukri_c3882_2021	G/T	2A	47.22	U	0.833758	0.566372	32/0	83/1	105/9	0/44	32/142	252/196	448
Excalibur_c12177_285	A/G	2A	47.22	U	0.833346	0.566396	0/32	1/83	9/104	44/0	142/32	196/251	447
Tdurum_contig30719_380	C/T	5A	12.95	S	0.820406	0.454956	0/32	62/23	115/0	44/0	173/1	394/56	450
BobWhite_c8266_227	G/T	5A	140.59	L	0.895911	0.596122	32/0	85/0	97/18	0/44	4/170	218/232	450
BS00063990_51	A/C	6A	60.27	S	0.713006	0.53887	32/0	82/3	102/13	2/42	51/123	269/181	450
Ex_c50864_326	A/G	6A	79.08	L	0.794176	0.593199	30/0	80/5	74/39	0/44	4/170	188/258	446
BS00065309_51	A/G	6A	79.08	U	0.765368	0.590065	30/2	80/3	74/40	0/44	4/168	188/257	445
wsnp_BE403818A_Ta_2_1	A/C	6A	84.66	L	0.783026	0.590404	0/32	7/78	67/48	44/0	174/0	292/158	450
Excalibur_rep_c105491_144	A/G	6A	84.66	U	0.782153	0.590547	0/32	7/78	66/49	44/0	174/0	291/159	450
IACX3586	A/G	6A	84.66	L	0.782153	0.590547	0/32	7/78	66/49	44/0	174/0	291/159	450
TA005615-0600	C/T	6A	84.66	U	0.782153	0.590547	0/32	7/78	66/49	44/0	174/0	291/159	450
IACX6046	C/T	6A	84.66	L	0.781383	0.590691	0/32	7/78	65/50	44/0	174/0	290/160	450
wsnp_Ex_rep_c67819_66516786	A/G	6A	85.01	L	0.787038	0.589886	0/32	7/78	70/44	44/0	174/0	295/154	449
Excalibur_c29707_318	A/G	1B	60.62	L	0.733209	0.566526	28/4	78/7	28/87	0/44	1/173	135/315	450
RAC875_c400_1363	C/T	1B	60.62	U	0.733209	0.566526	4/28	7/78	87/28	44/0	173/1	315/135	450
wsnp_Ex_c10233_16784994	C/T	1B	64.89	L	0.804438	0.589456	31/0	75/6	34/77	0/44	1/173	141/300	441
IAAV1869	A/G	1B	64.89	L	0.800804	0.590249	0/32	6/79	77/38	44/0	173/1	300/150	450
JD_c10376_670	A/G	1B	64.89	L	0.800804	0.590249	32/0	79/6	38/77	0/44	1/173	150/300	450
Kukri_c6462_320	C/T	1B	64.89	L	0.800804	0.590249	32/0	79/6	38/77	0/44	1/173	150/300	450
RAC875_rep_c108085_570	C/T	1B	64.89	U	0.800804	0.590249	0/32	6/79	77/38	43/0	173/1	299/150	449
RAC875_c21842_1647	A/G	1B	65.42	U	0.812444	0.591626	32/0	80/5	36/79	0/44	1/173	149/301	450
wsnp_Ku_c66585_65967792	C/T	1B	65.42	L	0.803572	0.590095	31/0	79/6	36/77	0/44	1/173	147/300	447
Kukri_rep_c95031_104	C/T	1B	65.42	L	0.800804	0.590249	32/0	79/6	38/77	0/44	1/173	150/300	450
RAC875_c24109_600	A/G	1B	65.42	L	0.800804	0.590249	0/32	6/79	77/38	44/0	173/1	300/150	450
Tdurum_contig11877_414	C/T	1B	65.42	L	0.800804	0.590249	0/32	6/79	77/38	44/0	173/1	300/150	450
Tdurum_contig65023_587	A/G	1B	65.42	L	0.800804	0.590249	0/32	6/79	77/38	44/0	173/1	300/150	450
Excalibur_c3140_697	A/G	1B	65.42	U	0.800781	0.590247	32/0	79/6	38/77	0/43	1/172	150/298	448
GENE-1214_159	A/G	1B	65.42	U	0.800781	0.590247	32/0	79/6	38/77	0/43	1/172	150/298	448
Kukri_c2297_181	C/T	1B	65.42	L	0.800781	0.590247	0/32	6/79	77/38	43/0	172/1	298/150	448
GENE-0116_102	C/T	1B	65.42	U	0.79917	0.589906	30/0	76/6	38/77	0/44	1/173	145/300	445
wsnp_Ex_c26419_35667216	A/G	1B	65.42	U	0.779469	0.590093	32/0	81/4	37/77	0/44	11/161	161/286	447

wsnp_Ku_c16117_24917524	A/G	1B	65.42	U	0.769735	0.589975	0/32	4/81	72/43	44/0	163/11	283/167	450
CAP11_c1969_268	A/C	1B	66.07	U	0.808293	0.591811	32/0	80/5	38/77	0/44	1/173	151/299	450
IACX184	C/T	1B	66.07	U	0.806373	0.591903	0/32	5/80	76/39	44/0	173/1	298/152	450
Ra_c18323_183	A/G	1B	66.07	U	0.806373	0.591903	0/32	5/80	76/39	44/0	173/1	298/152	450
Ex_c16691_96	A/G	1B	66.07	L	0.804954	0.591955	0/32	5/80	74/39	40/0	167/1	286/152	438
Kukri_c62285_336	A/G	1B	66.07	L	0.7995	0.592282	0/31	5/80	70/42	43/0	172/1	290/154	444
Kukri_rep_c88171_730	C/T	1B	66.07	L	0.800804	0.590249	0/32	6/79	77/38	44/0	173/1	300/150	450
BobWhite_c43322_203	A/C	1B	66.07	U	0.798319	0.592364	0/32	5/80	71/44	44/0	173/1	293/157	450
Tdurum_contig10326_151	A/G	1B	66.07	U	0.797017	0.592457	0/32	5/80	70/45	43/0	173/1	291/158	449
Tdurum_contig46780_203	A/G	1B	66.07	U	0.796994	0.592454	32/0	80/5	45/70	0/44	1/172	158/291	449
Tdurum_contig42856_1271	C/T	1B	66.07	U	0.796994	0.592454	32/0	80/5	45/70	0/42	1/172	158/289	447
BobWhite_c1318_691	C/T	1B	66.73	U	0.801377	0.586578	32/0	77/8	33/82	0/43	0/173	142/306	448
Kukri_c75399_389	A/G	1B	67.14	U	0.751917	0.556822	26/4	76/9	18/97	0/44	0/174	120/328	448
RAC875_c38384_450	C/T	1B	67.14	U	0.753141	0.558806	4/28	9/76	96/19	44/0	174/0	327/123	450
TA005233-0508	A/G	1B	67.14	U	0.801377	0.586578	0/32	8/77	82/33	43/0	173/0	306/142	448
Tdurum_contig42217_127	A/G	1B	67.14	L	0.812123	0.587367	0/32	6/79	92/23	44/0	165/9	307/143	450
Tdurum_contig43910_1113	A/G	1B	67.14	L	0.812123	0.587367	32/0	79/6	23/92	0/44	9/165	143/307	450
Tdurum_contig43910_687	A/G	1B	67.14	L	0.811585	0.587241	0/32	6/78	92/23	44/0	165/9	307/142	449
wsnp_BE442716B_Ta_2_2	A/G	1B	67.14	U	0.805214	0.587433	32/0	79/6	25/90	0/44	9/165	145/305	450
GENE-0235_131	A/G	1B	67.38	U	0.775062	0.590616	32/0	80/4	43/70	0/43	9/163	164/280	444
Excalibur_c1841_368	A/G	1B	68.04	L	0.838746	0.584614	0/32	8/77	94/21	44/0	174/0	320/130	450
Tdurum_contig25434_209	A/C	1B	68.04	U	0.819522	0.587876	32/0	79/6	24/91	0/44	6/168	141/309	450
RAC875_c24337_1280	C/T	1B	68.04	L	0.819823	0.587283	0/32	6/75	91/23	44/0	166/6	307/136	443
Kukri_c12151_769	A/G	1B	68.04	L	0.808642	0.587594	0/32	6/79	90/25	44/0	165/8	305/144	449
BobWhite_c11235_370	A/G	1B	68.04	U	0.805214	0.587433	32/0	79/6	25/90	0/44	9/165	145/305	450
BS00070139_51	A/C	1B	68.04	U	0.804477	0.58744	0/32	6/79	89/25	43/0	165/9	303/145	448
BS00071083_51	A/G	1B	68.04	L	0.805214	0.587433	32/0	79/6	25/90	0/44	9/165	145/305	450
BS00075663_51	A/G	1B	68.04	L	0.805214	0.587433	32/0	79/6	25/90	0/44	9/165	145/305	450
Kukri_c12151_467	A/G	1B	68.04	L	0.805214	0.587433	0/32	6/79	90/25	44/0	165/9	305/145	450
Tdurum_contig57731_162	C/T	1B	68.04	L	0.805214	0.587433	32/0	79/6	25/90	0/44	9/165	145/305	450
BS00038929_51	C/T	1B	68.04	L	0.805214	0.587433	32/0	79/6	25/90	0/43	9/165	145/304	449
Tdurum_contig57731_225	C/T	1B	68.04	L	0.804291	0.587431	0/32	6/79	89/25	44/0	164/9	303/145	448
BS00076696_51	C/T	1B	68.04	L	0.804291	0.587431	0/32	6/79	89/25	43/0	164/9	302/145	449
RAC875_c5796_424	A/G	1B	68.04	U	0.804291	0.587431	0/32	6/79	89/25	43/0	164/9	302/145	447
Tdurum_contig25434_218	A/C	1B	68.04	U	0.805028	0.587424	32/0	79/6	25/90	0/44	9/164	145/304	449
GENE-3085_533	A/G	1B	68.04	L	0.805028	0.587424	0/32	6/79	90/25	43/0	164/9	303/145	448
IACX5814	C/G	1B	68.04	U	0.804648	0.587406	32/0	79/6	25/90	0/43	9/162	145/301	446
Tdurum_contig57731_412	A/G	1B	68.04	L	0.80728	0.586619	0/28	6/72	90/24	44/0	165/8	305/132	437
RAC875_rep_c71685_445	A/G	1B	68.04	U	0.789759	0.587597	0/32	6/79	85/30	43/0	165/9	299/150	449
Excalibur_c59016_839	A/G	1B	68.04	U	0.789571	0.587585	0/32	6/79	85/30	43/0	164/9	298/150	448
RAC875_c6905_776	A/G	1B	68.04	U	0.789571	0.587585	32/0	79/6	30/85	0/43	9/164	150/298	448
RAC875_c32894_1116	A/G	1B	68.04	U	0.788996	0.58755	32/0	79/6	30/85	0/43	9/161	150/295	445
Ku_c10106_313	C/T	1B	70.08	U	0.79983	0.588287	0/32	7/78	80/35	44/0	173/1	304/146	450
Kukri_rep_c73393_848	A/G	1B	70.08	U	0.79983	0.588287	0/32	7/78	80/35	44/0	173/1	304/146	450
Excalibur_c29127_552	A/G	4B	64.58	L	0.808535	0.592327	32/0	80/5	40/75	0/43	0/171	152/294	446
tplb0056o05_409	A/C	4B	109.51	L	0.78928	0.591558	0/31	9/76	29/82	44/0	170/4	252/193	445

BS00062304_51	A/G	4B	110.84	L	0.83744	0.572	32/0	77/1	101/10	2/42	18/155	230/208	438
Excalibur_c43818_617	A/G	6B	115.25	L	0.80219	0.594866	32/0	81/4	50/60	0/44	0/174	163/282	445
BobWhite_c3515_753	A/G	6B	116.24	U	0.792037	0.589402	32/0	78/7	41/74	0/44	0/174	151/299	450
RAC875_c484_1063	A/G	6B	118.99	L	0.773217	0.562874	32/0	83/1	95/17	1/43	33/139	244/200	444
Tdurum_contig12648_389	C/T	6B	118.99	U	0.758812	0.556916	0/32	1/84	17/98	43/1	134/39	195/254	449
Tdurum_contig59339_428	C/T	6B	118.99	U	0.758812	0.556916	32/0	84/1	98/17	1/43	39/134	254/195	449
BS00012034_51	C/T	6B	118.99	U	0.759584	0.557456	0/31	1/73	16/93	43/1	133/38	193/236	429
Tdurum_contig59339_540	A/G	6B	118.99	U	0.757272	0.558016	30/0	80/1	93/17	1/43	38/134	242/195	437
BS00109878_51	C/T	6B	119.73	L	0.789225	0.582295	32/0	83/2	79/25	0/43	17/134	211/204	415
Kukri_c60966_261	A/G	6B	120.18	U	0.779529	0.562359	0/32	1/84	16/99	43/1	139/33	199/249	448
Tdurum_contig10729_986	A/G	6B	120.32	U	0.778215	0.563715	0/32	1/84	17/98	43/1	140/32	201/247	448
Tdurum_contig10729_734	C/T	6B	120.32	U	0.776515	0.563145	0/32	1/84	17/98	43/1	137/32	198/247	445
Kukri_c45876_157	A/G	6B	120.61	U	0.793772	0.589323	0/32	2/83	36/79	44/0	161/12	243/206	449
Tdurum_contig10729_989	C/T	6B	120.61	U	0.777653	0.563528	32/0	84/1	98/17	1/43	32/139	247/200	447
RAC875_c17011_717	A/G	6B	120.61	L	0.77483	0.562603	0/32	1/84	17/98	42/1	139/33	199/248	447
TA005876-0602	G/T	6B	120.61	L	0.743467	0.567508	32/0	83/2	90/25	1/43	31/142	237/212	449
TA002907-0816	C/T	6B	122.37	L	0.792346	0.591849	0/32	4/81	38/76	44/0	166/7	252/196	448
RAC875_c17011_373	C/T	6B	122.92	L	0.77201	0.591428	32/0	77/8	66/49	0/44	1/173	176/274	450
Kukri_c373_916	C/T	6B	122.92	U	0.773936	0.591407	32/0	80/5	53/59	0/44	5/168	170/276	446
RFL_Contig801_2124	C/T	7B	10.06	U	0.742037	0.568928	32/0	84/1	87/27	5/38	13/160	221/226	447
RAC875_c16839_188	C/T	7B	55.64	S	0.851078	0.58001	32/0	85/0	100/15	0/44	20/154	237/213	450
RAC875_c9715_679	C/T	7B	72.74	U	0.813325	0.590984	30/2	84/1	91/24	0/44	7/167	212/238	450
Tdurum_contig50984_553	A/G	7B	72.74	L	0.801997	0.590361	30/2	83/2	90/25	0/44	7/167	210/240	450
Tdurum_contig50984_599	A/G	7B	72.74	L	0.801255	0.590343	2/30	2/83	25/89	44/0	167/7	240/209	449
Kukri_c31305_75	A/C	7B	72.99	L	0.801255	0.590343	2/30	2/83	25/89	44/0	167/7	240/209	449
Ex_c6145_1877	G/T	1D	39.48	S	0.779088	0.59198	0/32	7/78	56/59	44/0	174/0	281/169	450
Ex_c6145_833	A/G	1D	39.48	S	0.779088	0.59198	0/32	7/78	56/59	44/0	174/0	281/169	450
RAC875_c7752_145	G/T	1D	39.48	S	0.779088	0.59198	32/0	78/7	59/56	0/44	0/174	169/281	450
RAC875_c7752_2913	A/G	1D	39.48	S	0.779088	0.59198	0/32	7/78	56/59	44/0	174/0	281/169	450
RAC875_c7752_549	A/G	1D	39.48	S	0.779088	0.59198	32/0	78/7	59/56	0/44	0/174	169/281	450
RAC875_c7752_1223	A/C	1D	39.48	S	0.779088	0.59198	32/0	78/7	59/56	0/44	0/173	169/280	449
Ku_c12842_664	A/G	1D	45.44	S	0.819371	0.596328	32/0	84/1	72/43	1/43	0/174	189/261	450
D_contig14507_369	A/G	1D	82.05	L	0.854169	0.593237	0/32	1/84	24/91	44/0	167/7	236/214	450
RFL_Contig1091_1538	C/T	5D	137.88	L	0.797123	0.589466	0/32	6/79	79/36	44/0	171/3	300/150	450
tplb0025h02_894	A/G	6D	133.61	L	0.772119	0.566941	0/32	2/83	18/97	42/2	149/25	211/239	450
tplb0025h02_630	C/T	6D	133.61	L	0.771344	0.566738	0/32	2/83	18/97	41/2	149/25	210/239	449
BobWhite_c12316_383	A/G	7D	26.33	S	0.790543	0.582968	0/32	6/79	21/94	43/1	155/12	225/218	443
Kukri_c80931_147	A/G	7D	26.33	S	0.794005	0.58067	32/0	79/5	94/19	1/40	12/129	218/193	411
BS00070188_51	G/T	7D	149.59	L	0.759203	0.523453	0/32	0/85	7/108	43/1	108/65	158/291	449
D_contig05962_325	G/T	7D	149.59	L	0.759203	0.523453	32/0	85/0	108/7	1/43	65/108	291/158	449
IAAV8855	C/T	7D	152.29	L	0.762516	0.525677	0/32	0/85	7/108	43/1	110/63	160/289	449
GENE-2323_270	A/C				0.842597	0.577494	32/0	74/11	15/100	0/44	0/174	121/329	450
BS00069075_51	A/C				0.857949	0.573785	0/32	0/85	10/104	44/0	149/25	203/246	449
BobWhite_c4126_442	A/G		unmapped		0.818089	0.578844	0/32	11/74	94/21	44/0	174/0	323/127	450
Ku_c11394_862	C/T				0.818358	0.575885	32/0	72/12	18/97	0/44	1/173	123/326	449
wsnp_Ku_c18023_27232712	A/C				0.798622	0.595169	0/32	4/81	52/63	44/0	173/1	273/177	450

Excalibur_c46262_263	A/G	0.792416	0.583413	32/0	79/6	94/21	1/43	12/162	218/232	450
TA001746-1415	A/G	0.792416	0.583413	0/32	6/79	21/94	43/1	162/12	232/218	450
RAC875_c2331_2570	C/T	0.797017	0.592457	0/32	5/80	70/45	44/0	173/1	292/158	450
CAP8_c950_198	C/T	0.792037	0.589402	32/0	78/7	41/74	0/44	0/174	151/299	450
IAAV2546	A/G	0.80244	0.578599	32/0	74/11	21/94	0/44	4/170	131/319	450
Kukri_c30982_1173	A/G	0.80244	0.578599	32/0	74/11	21/94	0/44	4/170	131/319	450
RAC875_c44575_396	A/G	0.80244	0.578599	32/0	74/11	21/94	0/44	4/170	131/319	450
wsnp_Ku_c30982_40765341	C/T	0.80244	0.578599	0/32	11/74	94/21	43/0	170/4	318/131	449
Kukri_c5631_770	C/T	0.786266	0.589974	0/32	7/78	70/45	44/0	174/0	295/155	450
BS00094584_51	A/C	0.778591	0.561179	32/0	84/1	96/15	1/43	34/138	247/197	444
BobWhite_c36693_210	C/T	0.743179	0.565547	32/0	83/2	92/23	2/42	28/146	237/213	450
Excalibur_c22696_316	A/C	0.736129	0.566914	0/32	19/66	85/30	44/0	173/1	321/129	450
Ku_c29409_324	A/G	0.736129	0.566914	0/32	19/66	85/30	44/0	173/1	321/129	450
GENE-4759_699	C/T	0.751575	0.554527	0/32	0/85	18/97	42/2	137/37	197/253	450
GENE-2128_156	C/T	0.717304	0.550281	0/32	27/58	90/25	44/0	174/0	335/115	450
Kukri_c51074_166	C/T	0.728648	0.54	32/0	84/1	86/27	1/40	34/131	237/199	436
Kukri_c27446_476	A/G	0.749375	0.529475	32/0	83/0	105/10	1/43	61/111	282/164	446
BobWhite_c31011_102	C/T	0.715648	0.536576	0/32	3/83	13/102	42/1	115/58	173/275	448
RAC875_rep_c110059_409	A/C	0.24058	0.09711	0/32	0/85	0/115	0/44	42/131	42/407	449
BS00083930_51	G/T	0.164265	0.066667	32/0	85/0	115/0	44/0	145/29	421/29	450

Strand: S = short arm, L = long arm, U = unknown

ExHet: Expected Heterozygosity

Table S6. Known QTL, Marker-Trait-Associations and MQTL co-localised with candidate loci for directional selection found

Chromosome	Position (cM)	Number of outlier loci	Co-localising locus	Position of co-localised locus (cM)	Effect on trait	Reference
1A	154.06	1	Xgwm1139, Xgwm750, Xbarc158.1	154.7	GY (minor)	Maccaferri et al. 2008
2A	25.97	2	QPro.inra-2A	20.6-37.5	GPS	http://ccg.murdoch.edu.au
			Lr17a	21.6	Leaf rust resistance gene	
	47.22	6	QHd.idw-2A.2	47	HD (major)	Maccaferri et al. 2008
5A	12.95	1	gwm443-wmc96	11.07	MQTL (function not described)	Zhang et al. 2010
	140.59	1	CAP7_c4064_162	137.88	HD increasing (major)	
			RAC875_c8642_231	141.75	HD increasing (major)	Zanke et al. 2014
			wsnp_Ex_c20899_30011827	147.26	HD increasing (major)	
			BobWhite_c8266_227*	140.59	Awn presence/absence	Mackay et al. 2014
			wsnp_Ex_c20899_30011827	147.26		
6A	60.27	1	QMxw.ucw-6A	43.5-71.6	Mixogram peak width	http://ccg.murdoch.edu.au
			QPsc.ucw-6A	43.5-61.1	Pasta color	
			QMxh.ucw-6A	43.5-58.5	Mixogram peak hight	
			Tdurum_contig50698_601	64.9	HD increasing (minor)	Zanke et al. 2014
	79.08	2				
	84.66-85.01	6	BS00037006_51	83.73	HD increasing (minor)	Zanke et al. 2014
1B	60.62	2	MQTL4	60.8	YLD, TGW, GPS, GWS, PHT (major)	Zhang et al. 2010
			QPsc.ucw-1B	57-79.1	Pasta color	http://ccg.murdoch.edu.au
			QPht.idw-1B.1	57.2-60.3	PHT	
			QTwt.ucw-1B	61.8-79.1	Test weight (milling yield)	
			QLd.sfr-1B	60.8-64.6	Lodging resistance	
			QLr.sfr-1B	62.1-64.6	Leaf rust resistance	
			QLtn.sfr.-1BS	63.3-65.9	Leaf rust reaction	
			YrH52	58.7	Yellow rust resistance	
			QEl.ipk-1B	64.6-79.8	Spike length	
			RFL_Contig529_963	60.62	HD decreasing (major)	Zanke et al. 2014
1B	64.89-68.04	58	BobWhite_rep_c66146_237	64.3	HD increasing (major)	
			BobWhite_c1318_691	66.73	HD increasing (minor)	
			D_contig17842_656	67.14	HD increasing (minor)	
			Excalibur_c15885_1145	64.32	HD increasing (minor)	
			IAAV3905	67.14	HD increasing (minor)	
			Ku_c31251_565	64.32	HD increasing (minor)	
			Tdurum_contig19102_84	64.3	HD increasing (minor)	
			Tdurum_contig31624_230	64.1	HD increasing (major)	
			Tdurum_contig83763_107	64.32	HD increasing (minor)	
1B	70.08	2	MQTL5	70.00	YLD, GPS, PHT, Tiller number (major)	Zhang et al. 2010

			BS00095568_51	70.08	HD decreasing (minor)	Zanke et al. 2014
			QGyld.agt-1B	72	YLD, Adaptation	Kuchel et al. 2007
			QFhs.whs-1B	71.6-76.6	FHB resistance	Klahr et al. 2007
4B	64.58	1	MQTL34	59	Rht-B1b	PHT, lodging resistance
			Excalibur_c56787_95	58.1	HD increasing (minor)	
4B	109.51-110.84	2	BobWhite_c47144_153	109.51	HD increasing (minor)	
			Tdurum_contig81113_395	112.16	HD decreasing (major)	Zanke et al. 2014
			Tdurum_contig81113_401	112.16	HD decreasing (major)	
			BS00100738_51	112.16	HD decreasing (major)	
6B	115.25-122.92	18	QSev.ucw-6B	103.7-119	SDS micro-sedimentation	http://ccg.murdoch.edu.au
			QLd.sfr-6B	89.9-109.8	Lodging resistance	
			Tdurum_contig92981_1091	113.67	HD decreasing (minor)	
			Excalibur_c32219_843	113.67	HD decreasing (minor)	
			36 markers mapped at this postion	118.99	HD decreasing (minor)	Zanke et al. 2014
			Tdurum_contig10729_986	120.32	HD decreasing (minor)	
			27 markers mapped at this postion	120.61	HD decreasing (minor)	
7B	10.06	1	QHd.idw-7B	5.5-15.5	HD (major)	Maccaferri et al. 2008
7B	55.64	1	Jagger_c9314_100	53.75	HD increasing (minor)	Zanke et al. 2014
7B	72.74-72.99	4	BS00053287_51	73.79	HD increasing (minor)	
1D	39.48	6	QRaw.ipk-1D.1	22.95-37.36	Awn color	
			QRg.ipk-1D.1	22.95-41.52	Glume color	http://ccg.murdoch.edu.au
			QRaw.ipk-1D.2	25.22-37.36	Awn color	
			7 markers mapped at this position	39.48	HD increasing (minor)	Zanke et al. 2014
1D	45.44	1	Sr33	42.6	Stem rust resistance	http://ccg.murdoch.edu.au
			Excalibur_c24303_1145	50.58	HD increasing (minor)	
1D	82.05	1	RAC875_c33279_526	81.93	HD increasing (major)	Zanke et al. 2014
			wsnp_Ku_c53270_57959459	81.93	HD increasing (major)	
5D	137.88	1				
6D	133.61	2	Ex_c7086_187	134.79	HD increasing (major)	Zanke et al. 2014
			RAC875_c63262_67	134.79	HD increasing (minor)	
7D	26.33	2				
7D	149.59-152.29	3			HD increasing (minor)	Zanke et al. 2014
					HD increasing (minor)	

Appendix II: Supplementary materials from:

Voss-Fels, K., Qian, L., Parra-Londono, S., Uptmoor, R., Frisch, M., Keeble-Gagnère, G., Appels, R., and Snowdon, R. J. (2016) Linkage drag constrains the roots of modern wheat. *Plant, Cell & Environment* (under review).

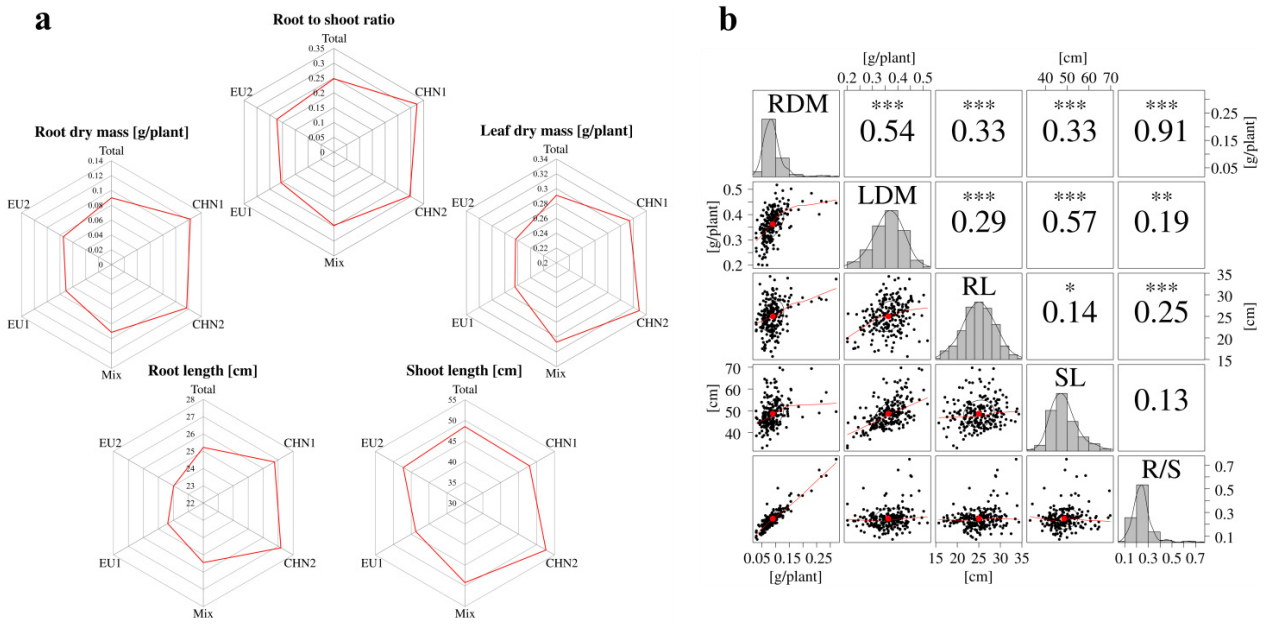


Fig. S1. Phenotypic variation among genetic subgroups and trait correlation matrix. (a) Mean values of the whole population and the five genetic subgroups among the 215 genotypes in the GWAS panel. (b) Pearson correlations among all traits for RDM = Rood dry mass, LDM = Leaf dry mass, RL = Root length, SL = Shoot length, R/S = Root to shoot ratio with significance thresholds * (p=0.05), ** (p=0.01), *** (p=0.001) and - (no significance); Scatterplots include LOESS fitting curves.

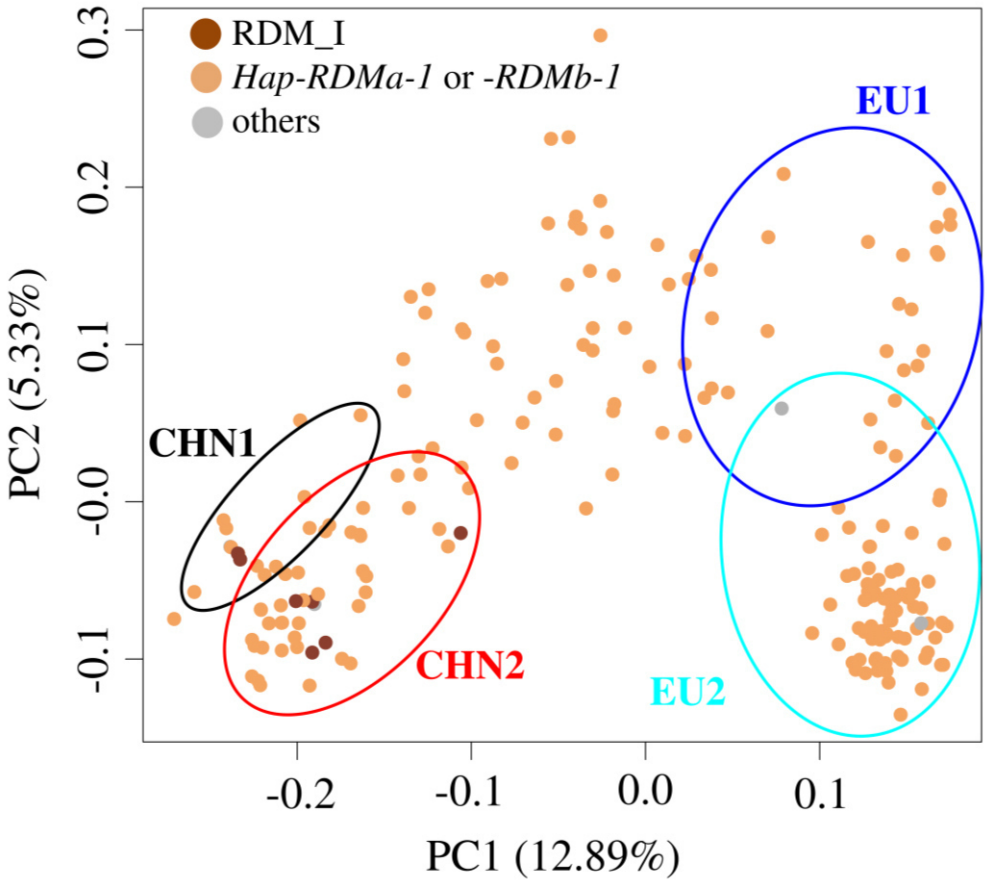


Fig. S2. Principal component analysis. Calculations are based on modified Roger's distances for 20,283 polymorphic single nucleotide polymorphism (SNP) markers (minor allele frequency ≥ 5%) and 215 genotypes. Colored ellipses represent the origins of the tested lines.

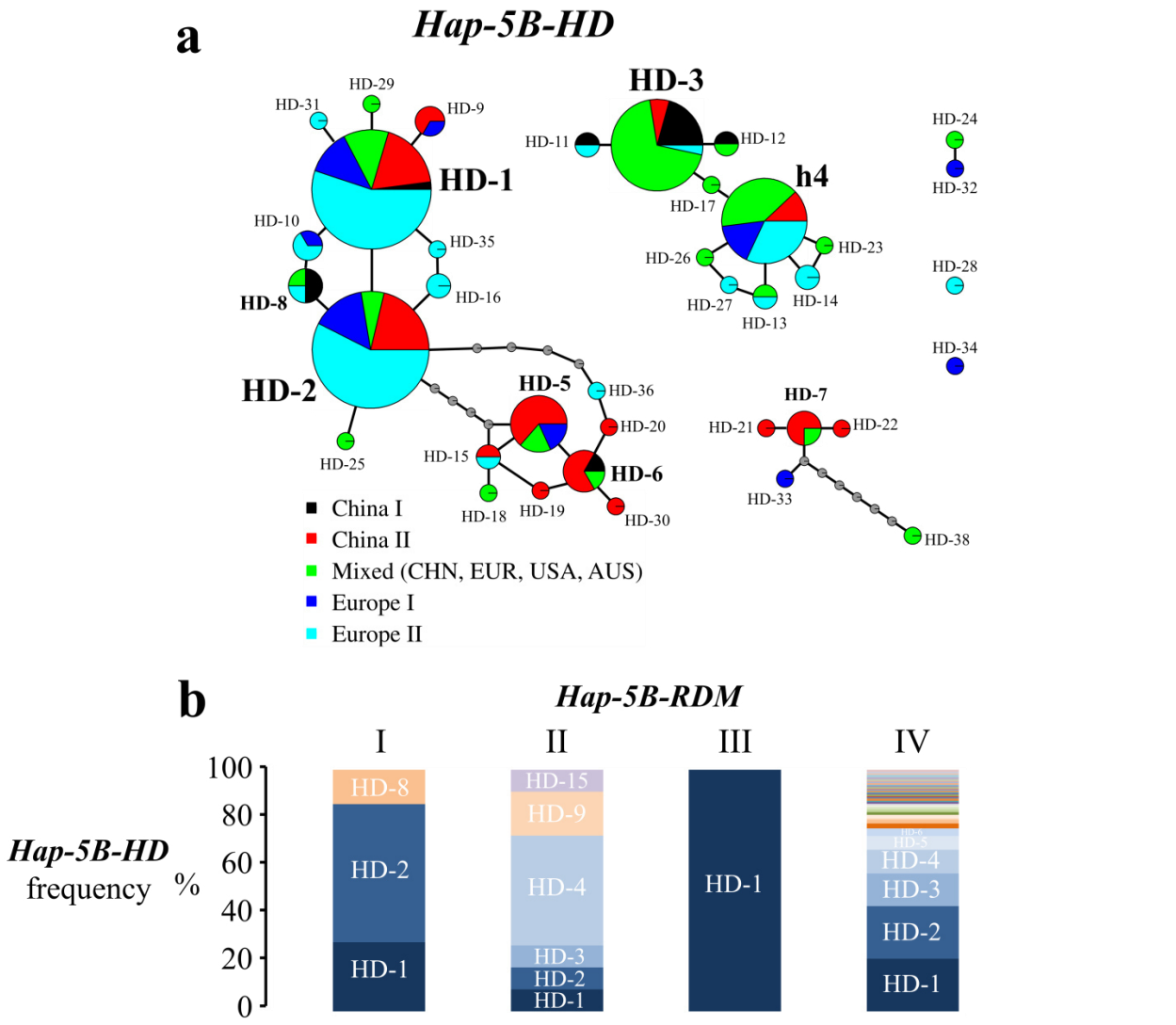


Fig S3. Haplotype network of the heading date (HD) related haplotype *Hap-5B-HD*. (a) The pie charts represent haplotype variants, their sizes are proportional to the number of genotypes that carry the respective variant and the colors indicate the gene pool affiliation. The networks give estimates on the genealogies of the haplotype variant sequences. All haplotype variants are differentiated from the nearest variant by one-nucleotide change in the haplotype sequence. Grey circles indicate additional probable sequence altering steps between two haplotype variations. (b) The colored bars represent haplotype variation frequencies for *Hap-5B-HD* in the respective RDM groups I (n=7), II (n=11), III (n=4), and IV (n=193).

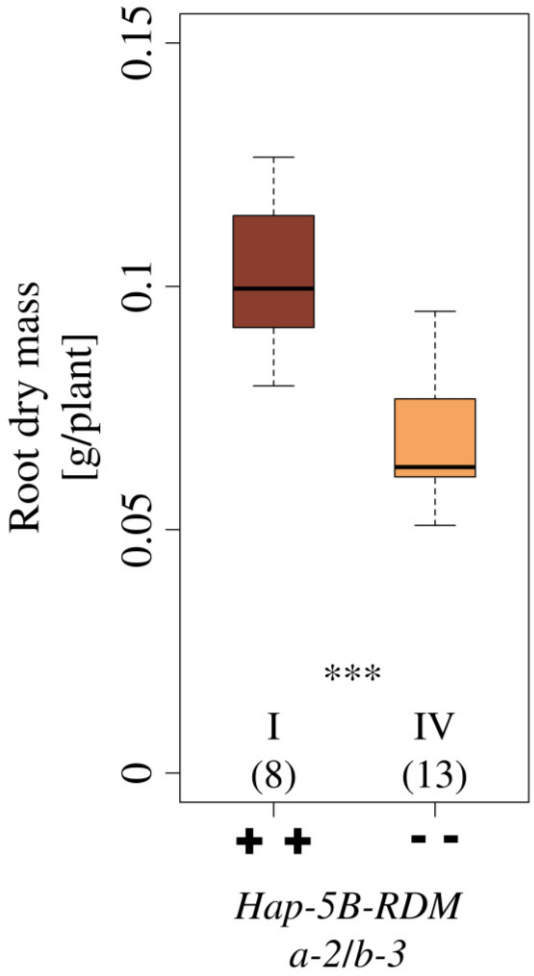


Fig S4. Phenotypic evaluation of the root dry mass (RDM) related haplotypes on RDM among 21 wheat lines. RDM was measured 35 days after growing with three replicates per genotype consisting of five plants per replication. RDM_I genotypes carry both favorable RDM haplotype variants *Hap-5B-RDMa-2/ Hap-5B-RDMb-3* while RDM_IV genotypes carry none of the positive variants. ANOVA significance threshold is *** (p=0.001).

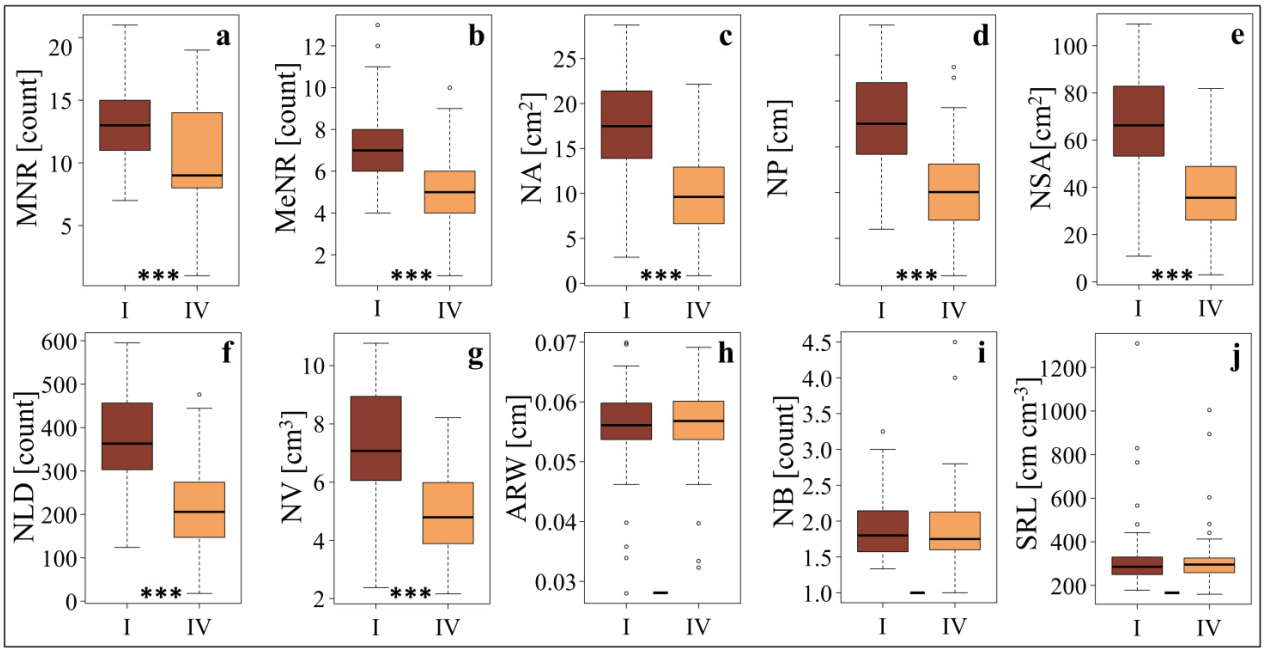


Fig. S5. Phenotypic evaluation of the root dry mass (RDM) related haplotypes on ten visual root traits among 24 wheat lines analysed with GiA roots software. Ten different traits were measured using the visual root analysis software GiA roots for seedling roots 23 days after sowing of which seven exceeded highly significant differences between RDM_I (n=15) and IV (n=9) genotypes (ANOVA significance threshold: *** p<0.001, - n.s.). (a) MNR= Maximum number of roots. (b) MeNR= Median number of roots. (c) NA= Network area. (d) NP= Network perimeter. (e) NSA= Network surface area. (f) NLD= Network length distribution. (g) NV= Network volume. (h) ARW= Average root width. (i) NB= Network bushiness. (j) SRL= Specific root length. Detailed trait description in Table S7.

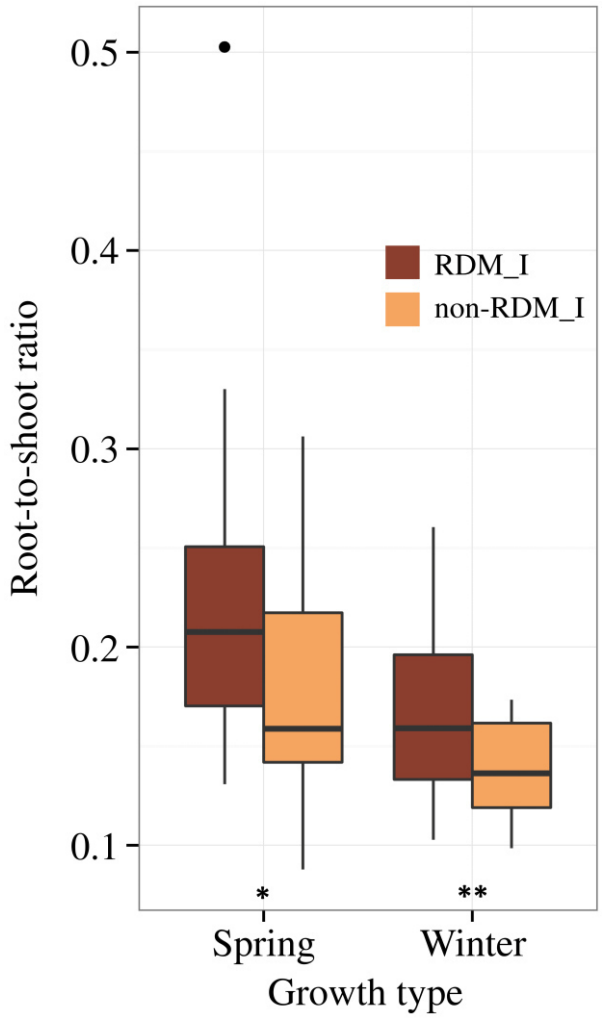


Fig. S6. Phenotypic evaluation of the root dry mass (RDM) related haplotypes on adult plants' RDM. (a) Root-to-shoot ratio (RS) was measured at BBCH stage 51-55 with three replicates per genotype consisting of four plants each. A total of 40 genotypes were analysed (20 genotypes per growth type group). RDM_I genotypes (seven for spring, eight for winter group) carry both favorable RDM haplotype variants *Hap-5B-RDMa-2*/ *Hap-5B-RDMb-3* while non-RDM_I genotypes do not carry this allelic combination. ANOVA significance threshold is * (p=0.05) and ** (p=0.01).

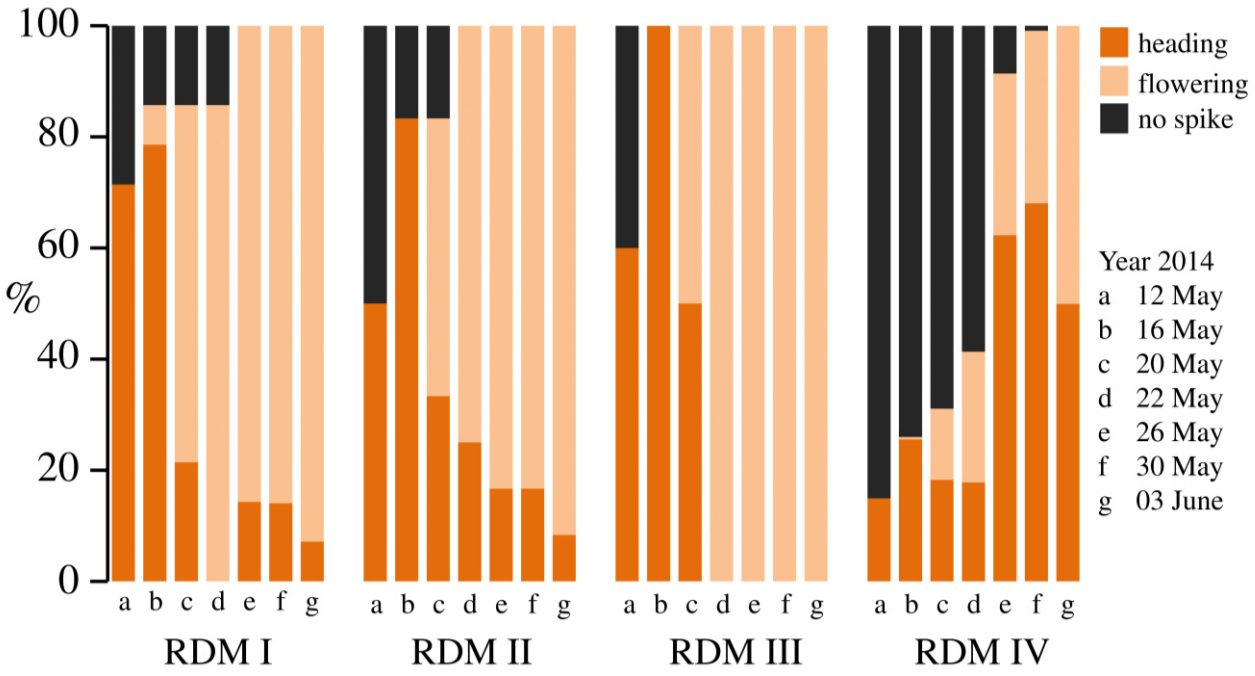


Fig. S7. Frequency of spike emergence and flowering time among 337 genotypes at seven different scoring time points. 183 genotypes of the 337 have been tested in at least one panel of this experiment. Plants were tested in field trials at Braunschweig/Germany in 2014 in two replicate plots per genotype. Number of genotypes tested: RDM_I = 7, RDM_II = 6, RDM_III = 5, RDM_IV = 319. Heading = BBCH 51-59, Flowering = BBCH 61-69).

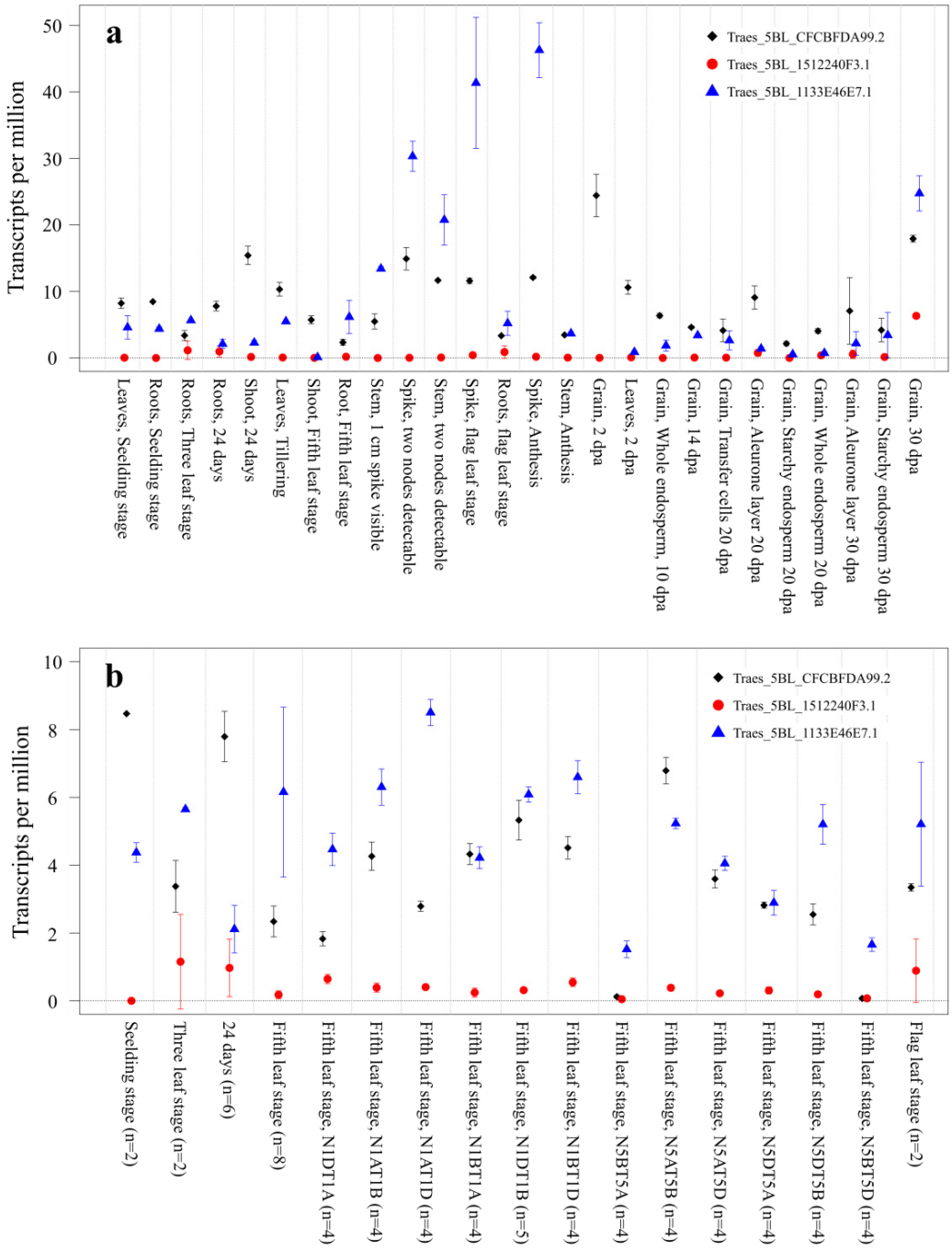


Fig. S8. Expression levels for selected genes. TPM (transcripts per million) expression levels downloaded from www.wheat-expression.com for Traes_5BL_CFCBFDA99.2, a cellulose synthase, Traes_5BL_1512240F3.1, an endo-beta-mannanase, and Traes_5BL_1133E46E7.1, an expansin in (a) wheat variety 'Chinese Spr' across different plant tissues and (b) in root tissue across all available varieties. The TPM mean value and the standard error are shown. dpa = days post anthesis. N1DT1A, N1AT1B, N1AT1D, N1BT1A, N1DT1B, N1BT1D, N5BT5A, N5AT5B, N5AT5D, N5DT5A, N5DT5A, N5DT5A and N5BT5D are Nulli-Tetra lines of 'Chinese Spr' that miss a given chromosome. Two additional copies of another chromosome compensate for the absence of the respective chromosome; e.g. N5BT5D has four copies of 5D that compensate for the absence of 5B.

Table S1. Genotype summaryof the tested wheat lines

ID-No.	Genotype name	Subgroup assignment based on k-means clustering	Growth type	Time of flowering	Origin	Pedigree (if available)	In GWAS panel?	In evaluation panel I?	In evaluation panel II (root scanning)?	In evaluation panel III (adult plants)?	In heading date scoring panel?
SW070	WvB/W-SW070	1	spr	early	China		x				
SW097	WvB/W-SW097	1	spr	early	China		x				
SW098	WvB/W-SW098	1	spr	early	China		x				
SW100	WvB/W-SW100	1	spr	early	China		x				
WW103	Shengxuan4	1	win	early	China		x				x
SW107	WvB/W-SW107	1	spr	early	China		x	x	x	x	
SW110	WvB/W-SW110	1	spr	early	China		x				
SW150	WvB/W-SW150	1	spr	early	China		x				
SW156	WvB/W-SW156	1	spr	early	China		x				
WW374	Ning05562	1	win	early	China		x				x
WW381	Ning0569	1	win	early	China		x				x
WW382	WvB/W-WW382	1	win	early	China		x	x	x		x
SW905	Wangshubai	1	spr	early	China		x				
WW002	WvB/W-WW002	2	win	early	China		x				x
WW003	WvB/W-WW003	2	win	early	China		x				x
WW19	WvB/W-WW19	2	win	early	China		x				x
WW020	WvB/W-WW020	2	win	early	China		x				x
SW063	WvB/W-SW063	2	spr	early	China		x				
WW069	WvB/W-WW069	2	win	early	China		x				x
SW126	WvB/W-SW126	2	spr	early	China		x	x	x	x	
WW134	Yangmai17	2	win	early	China		x				x
WW135	Ningyan1	2	win	early	China		x				x
WW136	Ningnuo1	2	win	early	China		x				x
WW137	Yangmai11	2	win	early	China		x				x
WW138	WvB/W-WW138	2	win	early	China		x				x
WW139	Yangmai13	2	win	early	China		x				x
WW140	Yangmai14	2	win	early	China		x				x
SW149	WvB/W-SW149	2	spr	early	China		x				
WW151	WvB/W-WW151	2	win	early	China		x				x
WW154	WvB/W-WW154	2	win	early	China		x				x
SW159	WvB/W-SW159	2	spr	early	China		x				
WW160	WvB/W-WW160	2	win	early	China		x				x
WW303	W-WW303	2	win	late	Netherlands	TABASCO x SHA3/CBRD	x				x
WW365	WvB/W-WW365	2	win	early	China		x				x

SW371	Xumai29	2	spr	early	China	x				
WW375	Ning0762	2	win	early	China	x				x
WW376	Ning0644	2	win	early	China	x				x
WW378	Ning0604	2	win	early	China	x	x	x		x
WW379	Ning0588	2	win	early	China	x				x
WW383	Ningmai14	2	win	early	China	x				x
SW386	WvB/W-SW386	2	spr	early	China	x				
SW387	Yang 06G5	2	spr	early	China	x				
SW388	WvB/W-SW388	2	spr	early	China	x	x	x	x	
WW392	WvB/W-WW392	2	win	early	China	x	x	x		x
SW396	WvB/W-SW396	2	spr	early	China	x				
SW402	WvB/W-SW402	2	spr	early	China	x				
WW405	WvB/W-WW405	2	win	early	China	x				x
WW406	WvB/W-WW406	2	win	early	China	x				x
SW408	WvB/W-SW408	2	spr	early	China	x				
SW414	WvB/W-SW414	2	spr	early	China	x				
WW415	Sumai 2	2	win	early	China	x				x
WW417	WvB/W-WW417	2	win	early	China	x				x
WW419	WvB/W-WW419	2	win	early	China	x	x	x		x
WW422	WvB/W-WW422	2	win	early	China	x				x
WW423	WvB/W-WW423	2	win	early	China	x				x
WW424	WvB/W-WW424	2	win	early	China	x				x
WW430	WvB/W-WW430	2	win	early	China	x				x
SW536	WvB/W-SW536	2	spr	early	China	x				
SW538	WvB/W-SW538	2	spr	early	China	x			x	
WW007	WvB/W-WW007	3	win	early	China	x				x
WW13	WvB/W-WW13	3	win	early	China	x				x
WW15	WvB/W-WW15	3	win	early	China	x				x
WW16	WvB/W-WW16	3	win	early	China	x				x
WW059	WvB/W-WW059	3	win	early	North America	x				x
WW061	WvB/W-WW061	3	win	late	Slovakia	x				x
SW074	Sumai1	3	spr	early	China	x				
SW108	WvB/W-SW108	3	spr	early	China	x				
SW109	WvB/W-SW109	3	spr	early	China	x				
SW112	WvB/W-SW112	3	spr	early	China	x			x	
SW114	WvB/W-SW114	3	spr	early	China	x				
SW162	WvB/W-SW162	3	spr	early	China	x				
FWW166	Han6172	3	fac win	early	China	x				x
FSW167	Zhoumai16	3	fac spr	early	China	x				
FSW170	Jimai20	3	fac win	early	China	x				
FSW172	Zhengmai366	3	fac spr	early	China	x				

FSW174	Ji3475	3	fac spr	early	China		x			
FSW175	Kaocheng8901	3	fac spr	early	China		x		x	
FSW176	Gaoyou503	3	fac spr	early	China		x			
FWW178	Zhong6	3	fac win	early	China		x			x
FWW180	Yumai35	3	fac win	early	China		x			x
FSW184	Zhoumai13	3	fac spr	early	China		x			
FWW185	WvB/W-FWW185	3	fac win	early	China		x		x	x
FWW186	Lankao24	3	fac win	early	China		x			x
FWW187	WvB/W-FWW187	3	fac win	early	China		x			x
WW201	WvB/W-WW201	3	win	early	China		x			x
WW202	Jingdong17	3	win	early	China		x			x
WW203	Jinmai61	3	win	early	China		x			x
WW204	Jinong207	3	win	early	China		x			x
WW224	Intro	3	win	late	Germany		x			x
WW293	W-WW293	3	win	late	Netherlands	W x NG8675/CBRD	x			x
WW294	W-WW294	3	win	late	Netherlands	W x NG8675/CBRD	x			x
WW296	W-WW296	3	win	late	Netherlands	W x NG8675/CBRD	x			x
WW297	W-WW297	3	win	late	Netherlands	W x NG8675/CBRD	x			x
WW298	W-WW298	3	win	late	Netherlands	W x NG8675/CBRD	x			x
WW299	W-WW299	3	win	late	Netherlands	W x SHA3/CBRD	x			x
SW300	W-SW300	3	spr	late	Netherlands	W x SHA3/CBRD	x		x	x
WW304	W-WW304	3	win	late	Netherlands	TABASCO x SHA3/CBRD	x			x
WW306	W-WW306	3	win	late	Netherlands	TABASCO x SHA3/CBRD	x			x
SW309	W-SW309	3	spr	late	Netherlands	TABASCO x SHA3/CBRD	x		x	x
WW313	W-WW313	3	win	late	Netherlands	TABASCO x SHA3/CBRD	x			x
WW315	W-WW315	3	win	late	Netherlands	TABASCO x SHA3/CBRD	x			x
WW363	WvB/W-WW363	3	win	early	China		x			x
WW364	WvB/W-WW364	3	win	early	China		x			x
WW369	Huaimai 21	3	win	early	China		x			
SW372	Xumai27	3	spr	early	China		x			
SW373	Ning07233	3	spr	early	China		x			
SW389	Mai 48	3	spr	early	China		x			
SW394	WvB/W-SW394	3	spr	early	China		x			
WW418	WvB/W-WW418	3	win	early	China		x			x
SW426	J95	3	spr	early	China		x			
WW545	JLU-WW545	3	win	late	Germany		x			x
SW902	Florence Aurore	3	spr	early	France		x		x	
SW904	WvB/W-SW904	3	spr	early	China		x			
SW906	CSCR6	3	spr	early	Australia		x			
SW907	CSCR14	3	spr	early	Australia		x			
SW908	CSCR16	3	spr	early	Australia		x			
SW909	CSCR28	3	spr	early	Australia		x			
WW310	W-WW310	4	win	late	Netherlands	TABASCO x SHA3/CBRD	x			x
WW320	W-WW320	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW322	W-WW322	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW325	W-WW325	4	win	late	Netherlands	W x SHA3/CBRD	x	x		x
WW326	W-WW326	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW328	W-WW328	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW329	W-WW329	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW331	W-WW331	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW332	W-WW332	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW335	W-WW335	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW338	W-WW338	4	win	late	Netherlands	W x SHA3/CBRD	x			x
WW553	WvB/W-	4	win	late	Germany		x			x

	WW553								
WW562	WvB/W-WW562	4	win	late	Germany		x		x
WW576	WvB/W-WW576	4	win	late	Germany		x		x
WW599	WvB/W-WW599	4	win	late	Germany		x		x
WW603	WvB/W-WW603	4	win	late	Germany		x		x
WW604	WvB/W-WW604	4	win	late	Germany		x		x
WW609	WvB/W-WW609	4	win	late	Germany		x		x
WW633	WvB/W-WW633	4	win	late	Germany		x		x
WW004	WvB/W-WW004	5	win	early	China		x		x
WW006	WvB/W-WW006	5	win	early	China		x		x
WW011	Tobak	5	win	late	Germany	(Ellvis x Drifter) x Koch	x		x
WW12	WvB/W-WW12	5	win	early	China		x		x
WW18	WvB/W-WW18	5	win	early	China		x		x
WW060	WvB-WW060	5	win	late	Romania		x		x
WW199	Yunmai46	5	win	early	China		x		x
WW205	WvB-WW205	5	win	late	Germany	(Ilias x Darwin) x Koch	x		x
WW206	WvB-WW206	5	win	late	Germany	Tyberius x Opus	x		x
WW207	WvB-WW207	5	win	late	Germany	(Qualibo x Tommi) x Tulsa	x		x
WW208	WvB-WW208	5	win	late	Germany	(China 1 x Opus) x Tulsa	x		x
WW210	Alchemy	5	win	late	France		x		x
WW218	Elixer	5	win	late	Germany	(Semper x Bristol) x Tulsa	x		x
WW219	Esket	5	win	late	Germany		x		x
WW220	Genius	5	win	late	Germany		x		x
WW221	Gladiator	5	win	late	UK		x		x
WW222	Global	5	win	late	Germany		x		x
WW223	Inspiration	5	win	late	Germany		x		x
WW225	JB Asano	5	win	late	Germany		x		x
WW226	Julius	5	win	late	Germany		x		x
WW227	Kerubino	5	win	late	Germany		x		x
WW229	Kredo	5	win	late	Germany		x		x
WW234	Lear	5	win	late	Germany		x		x
WW236	Lucius	5	win	late	Germany		x		x
WW238	Muskat	5	win	late	Germany	ZE 21372 x 86Z99-9	x		x
WW244	Potenzial	5	win	late	Germany		x		x
WW253	Tabasco	5	win	late	Germany	(ZE.90-2666 x 86Z99.9) x CPB.93-27	x		x
WW267	Pionier	5	win	late	Germany		x		x
WW271	Edward	5	win	late	Germany	(Corvus x Bristol) x Biscay	x		x
WW288	BB 619609	5	win	late	Germany	Hermann x Tulsa	x		x
WW289	BB 690909	5	win	late	Germany	(Qualibo x Tommi) x Tulsa	x		x
WW292	W-WW292	5	win	late	Netherlands	W x NG8675/CBRD	x	x	x
WW314	W-WW314	5	win	late	Netherlands	TABASCO x SHA3/CBRD	x		x
WW547	WvB/JLU-WW547	5	win	late	Germany		x		x
WW548	WvB/W-WW548	5	win	late	Germany		x		x
WW549	Accroc	5	win	late	France		x		x
WW550	Alixan	5	win	late	France		x		x
WW551	Sachsmo	5	win	late	Germany		x		x
WW552	WvB/W-WW552	5	win	late	Germany		x		x
WW556	WvB/W-WW556	5	win	late	Germany		x		x
WW558	WvB/W-WW558	5	win	late	Germany		x		x
WW559	WvB/W-WW559	5	win	late	Germany		x		x
WW564	WvB/W-WW564	5	win	late	Germany		x		x
WW567	WvB/W-	5	win	late	Germany		x		x

	WW567								
WW570	WvB/W-WW570	5	win	late	Germany	x			x
WW578	WvB/W-WW578	5	win	late	Germany	x			x
WW581	WvB/W-WW581	5	win	late	Germany	x			x
WW584	WvB/W-WW584	5	win	late	Germany	x			x
WW590	WvB/W-WW590	5	win	late	Germany	x			x
WW592	WvB/W-WW592	5	win	late	Germany	x			x
WW593	WvB/W-WW593	5	win	late	Germany	x			x
WW594	WvB/W-WW594	5	win	late	Germany	x			x
WW595	WvB/W-WW595	5	win	late	Germany	x			x
WW597	WvB/W-WW597	5	win	late	Germany	x			x
WW598	WvB-WW598	5	win	late	Germany	x			x
WW601	WvB/W-WW601	5	win	late	Germany	x			x
WW607	WvB/W-WW607	5	win	late	Germany	x			x
WW608	WvB/W-WW608	5	win	late	Germany	x			x
WW611	WvB/W-WW611	5	win	late	Germany	x			x
WW613	WvB/W-WW613	5	win	late	Germany	x			x
WW614	WvB/W-WW614	5	win	late	Germany	x			x
WW615	WvB/W-WW615	5	win	late	Germany	x			x
WW620	WvB/W-WW620	5	win	late	Germany	x			x
WW623	WvB/W-WW623	5	win	late	Germany	x			x
WW625	WvB/W-WW625	5	win	late	Germany	x			x
WW626	WvB/W-WW626	5	win	late	Germany	x			x
WW629	WvB/W-WW629	5	win	late	Germany	x			x
WW631	KWS Milaneco	5	win	late	Germany	x			x
WW634	WvB/W-WW634	5	win	late	Germany	x			x
WW636	WvB/W-WW636	5	win	late	Germany	x			x
WW638	WvB/W-WW638	5	win	late	Germany	x			x
WW640	WvB/W-WW640	5	win	late	Germany	x			x
WW644	WvB/W-WW644	5	win	late	Germany	x			x
WW646	WvB/W-WW646	5	win	late	Germany	x			x
WW648	WvB/W-WW648	5	win	late	Germany	x			x
WW649	WvB-WW649	5	win	late	Germany	x			x
WW900	WvB-WW900	5	win	late	Germany	x			
WW901	Dream	5	win	late	Germany	x			
WW903	Lynx	5	win	late	UK	x			
SW155	WvB/W-SW155	1	spr	early	China		x	x	x
WW029	WvB/W-WW029	2	win	early	China		x		x
WW033	WvB/W-WW033	2	win	early	China		x	x	x

SW91	Ningmai 12	2	spr	early	China		x		x	
SW099	WvB/W-SW099	2	spr	early	China		x	x	x	
SW161	WvB/W-SW161	2	spr	early	China		x		x	
WW311	W-WW311	2	win	late	Netherlands	TABASCO x SHA3/CBRD	x	x		x
WW380	Ning0575	2	win	early	China		x			x
SW384	WvB/W-SW384	2	spr	early	China		x	x	x	
WW393	WvB/W-WW393	2	win	early	China		x	x		x
SW398	WvB/W-SW398	2	spr	early	China		x	x	x	
SW401	WvB/W-SW401	2	spr	early	China		x		x	
SW411	Shibin14	2	spr	early	China		x	x	x	
SW065	WvB/W-SW065	3	spr	early	China				x	
WW131	WvB/W-WW131	3	win	early	China		x			x
SW420	WvB/W-SW420	3	spr	early	China		x		x	
WW327	W-WW327	4	win	late	Netherlands	W x SHA3/CBRD	x	x		x
WW339	W-WW339	4	win	late	Netherlands	W x SHA3/CBRD	x	x		x
WW212	Beluga	5	win	late	UK		x			x
WW241	Orcas	5	win	late	Germany		x	x		x
WW245	Premio	5	win	late	Germany		x			x
WW250	Stigg	5	win	late	UK		x			x
WW255	Warrior	5	win	late	UK		x			x
WW269	Gordian	5	win	late	Germany		x	x		x
WW287	WvB-WW287	5	win	late	Germany	(Roswell x Privileg) x Bristol	x	x		x
WW579	WvB/W-WW579	5	win	late	Germany		x			x
WW582	WvB/W-WW582	5	win	late	Germany		x	x		x
WW586	WvB/W-WW586	5	win	late	Germany		x	x		x
WW596	WvB/W-WW596	5	win	late	Germany		x			x

Table S2: Summary of adjusted phenotype mean values.

		RDM [g/plant]	LDM [g/plant]	RL [cm]	SL [cm]	R/S
Mean	Total	0.090	0.362	25.01	48.63	0.247
	China I	0.123	0.379	26.76	47.93	0.324
	China II	0.117	0.395	27.15	51.87	0.296
	Mixed	0.091	0.373	25.47	49.25	0.247
	Europe I	0.070	0.341	24.64	44.09	0.206
	Europe II	0.075	0.340	23.34	47.85	0.222
Minimum	Total	0.029	0.201	15.70	32.92	0.073
	China I	0.067	0.289	21.036	43.849	0.183
	China II	0.064	0.302	20.30	41.83	0.163
	Mixed	0.033	0.201	15.70	32.92	0.073
	Europe I	0.029	0.225	16.86	37.93	0.116
	Europe II	0.029	0.201	16.86	33.94	0.089
Maximum	Total	0.321	0.517	34.29	69.76	0.750
	China I	0.259	0.432	29.99	51.85	0.659
	China II	0.321	0.502	34.29	69.38	0.750
	Mixed	0.267	0.517	33.45	69.76	0.603
	Europe I	0.111	0.429	31.32	52.11	0.278
	Europe II	0.143	0.464	31.02	64.01	0.372
SD	Total	0.040	0.059	3.58	6.29	0.088
	China I	0.057	0.042	2.57	3.07	0.133
	China II	0.052	0.042	3.26	5.51	0.110
	Mixed	0.035	0.061	3.54	7.53	0.085
	Europe I	0.020	0.053	3.51	3.28	0.048
	Europe II	0.022	0.058	3.14	5.85	0.056

RDM = Root dry mass, LDM = Leaf dry mass, RL = Root length, SL = Shoot length, R/S = root-to-shoot ratio

Table S3: Summary of significant marker-trait-associations.

Trait	SNP	Chromosome	Position (cM)	–log10(p-value)	Allele	Effect on trait
RDM [g/plant]	Tdurum_contig48959_1172	5B	137.1	5.7	T	0.0229228
	GENE-2890_482		143.55	5.92	A	0.02008483
	Excalibur_c25522_755		143.55	5.91	T	0.02005586
	Kukri_c46570_214		143.55	5.88	G	0.02005586
	RAC875_c12293_588		143.55	5.87	A	0.02005586
	RAC875_c18088_2222		143.55	5.82	A	0.02005586
	BobWhite_c43_86		143.55	5.8	A	0.02003887
	RAC875_c18088_950		143.55	5.79	T	0.02001383
	RAC875_c24226_1356		143.55	5.77	C	0.01998779
	Excalibur_c60554_394		143.55	5.76	T	0.01995988
RL [cm]	BS00076003_51	2B	130.62	4.29	T	-1.2142134
	GENE-1280_188		130.62	4.19	C	-1.1976762
	IAAV1650	5A	89.56	4.3	G	-1.3907469
	Excalibur_rep_c103747_193		89.56	4.28	C	-1.3860799
	wsnp_Ex_c13942_21820758		89.56	4.24	T	-1.3806169
SL [cm]	RAC875_c98242_422	6D	22.92	4.33	G	2.821723

RDM = Root dry mass, LDM = Leaf dry mass, RL = Root length, SL = Shoot length, R/S = root-to-shoot ratio; SNPs with $-\log_{10}$(p-value) > 4 were considered significant.

Table S4: Haplotype variants for *Hap-5B-RDMa* and the different variants within the genotype collection

		GENE-2890_482	Excalibur_c25522_755	Kukri_c46570_214	RAC875_c12293_588	RAC875_c18088_2222	BobWhite_c43_86	RAC875_c18088_950	RAC875_c24226_1356	Excalibur_c60554_394	Counts					
	Alleles	G/A	C/T	A/G	G/A	G/A	G/A	C/T	A/C	G/T	China I	China II	Mixed	Europe I	Europe II	**Total**
Hap-5B-RDMa	h1	G	C	A	G	G	G	C	A	G	8	32	49	24	73	**186**
	h2	A	T	G	A	A	A	T	C	T	2	11	3	-	2	**18**
	h3	G	C	A	G	G	G	-	A	G	-	1	1	-	2	**4**
	h4	G	C	A	G	G	G	C	-	G	-	2	-	1	-	**3**
	h5	G	C	A	G	G	G	C	A	-	1	-	-	-	1	**2**
	h6	G	C	A	G	G	-	C	A	G	-	-	1	-	-	**1**
	h7	-	C	A	G	G	G	-	-	-	-	-	-	-	1	**1**
	Number of genotypes										**11**	**46**	**54**	**25**	**79**	**215**

Significant SNPs from GWAS are marked in red

Table S5: Haplotype variants for *Hap-5B-RDMb* and the different variants within the genotype collection

		BS00022231_51	BS00022477_51	BS00029852_51	BS00110293_51	IACX6288	Tdurum_contig48959_1172	Counts					
	Alleles	G/A	G/A	G/T	G/A	G/A	G/T	China I	China II	Mixed	Europe I	Europe II	**Total**
Hap-5B-RDMb	h1	G	A	G	A	T	G	1	31	41	20	67	**160**
	h2	A	G	T	G	C	G	7	4	8	4	6	**29**
	h3	G	G	T	G	C	T	2	9	-	-	-	**11**
	h4	G	A	-	A	T	G	-	1	3	1	4	**9**
	h5	A	G	T	G	-	G	1	-	-	-	1	**2**
	h6	G	G	T	G	-	T	-	1	-	-	-	**1**
	h7	G	A	G	A	-	G	-	-	1	-	-	**1**
	h8	G	A	G	G	C	T	-	-	1	-	-	**1**
	h9	G	-	-	A	T	G	-	-	-	-	1	**1**
	Number of genotypes							**11**	**46**	**54**	**25**	**79**	**215**

Significant SNPs from GWAS are marked in red

Table S6: Haplotype variants for *Hap-5B-HD* and the different variants within the genotype collection

Hap-5B-HD	Alleles	BobWhite_c11038_605	BS00037023_51	BobWhite_c14360_420	BS00021868_51	BS00064272_51	BS00064274_51	BS00084106_51	BS00093743_51	BS00093746_51	Excalibur_c23661_1712	Excalibur_c32979_1152	Excalibur_c3328_227	Excalibur_c39672_189	Excalibur_c39672_90	Excalibur_c4083_731	GENE-2855_121	IAAV4395	IAAV5779	IACX6116	IACX6772	IACX7841	Kukri_c11272_486	Kukri_c16006_1996	Kukri_c16006_431	Kukri_c19178_2327	Kukri_rep_c103049_96	RAC875_c38276_2128	RAC875_rep_c105761_317	wsnp_Ex_c23661_32900048	wsnp_JD_c5431_6580242	wsnp_Ku_c40084_48381107	wsnp_RFL_Contig3139_3096141	wsnp_RFL_Contig3238_3265410	BobWhite_c38408_71	BS00050054_51	BS00050057_51	BS00097105_51	Kukri_rep_c112425_506	Kukri_rep_c112425_98	RAC875_c6890_150	Tdurum_contig8171_1548	Tdurum_contig8171_1602	Tdurum_contig8171_1680	Tdurum_contig8171_1712	Tdurum_contig8171_1772	Tdurum_contig8171_1783	Counts					
		C/T	C/T	T/C	G/A	G/A	G/A	A/G	T/C	C/T	G/T	A/G	T/C	G/T	C/T	G/T	C/T	C/T	C/T	A/G	T/G	T/C	C/T	A/G	T/C	C/T	A/G	A/G	G/A	G/A	A/G	T/G	G/A	T/C	C/T	T/C	G/T	A/G	A/G	T/C	C/T	C/A	T/C	A/G	T/C	A/C	T/C	China I	China II	Mixed	Europe I	Europe II	**Total**
	1	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	1	9	6	6	27	**49**
	2	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	10	3	7	27	**47**
	3	T	T	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	C	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	6	2	20	0	1	**29**
	4	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	C	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	3	10	4	8	**25**
	5	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	T	C	C	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	7	2	2	-	**11**
	6	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	T	C	C	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	1	4	1	-	-	**6**
	7	T	T	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	C	C	G	T	T	G	A	G	G	G	T	G	T	C	T	T	G	G	C	T	A	C	G	C	A	C	-	3	1	-	-	**4**
	8	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	-	A	C	1	-	1	1	1	**4**
	9	C	T	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	2	-	1	-	**3**
	10	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	-	A	C	-	-	-	1	2	**3**
	11	T	T	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	-	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	1	-	-	-	1	**2**
	12	T	T	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	C	G	T	T	G	-	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	1	-	1	-	-	**2**
	13	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	C	G	T	T	G	-	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	1	-	1	**2**
	14	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	-	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	-	-	2	**2**
	15	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	T	C	-	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	1	-	-	1	**2**
	16	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	-	T	G	G	C	T	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	-	-	2	**2**
	17	T	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	C	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	1	-	-	**1**
	18	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	T	C	-	A	C	-	A	G	A	A	A	G	A	C	C	T	T	-	G	C	T	A	C	G	C	A	C	-	-	1	-	-	**1**
	19	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	T	C	-	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	1	-	-	-	**1**

20	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	-	C	C	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	1	-	-	-	**1**
21	T	T	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	C	C	G	T	T	G	-	G	G	G	T	G	T	C	T	T	G	G	C	T	A	C	G	C	A	C	-	1	-	-	-	**1**
22	T	T	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	C	C	G	T	T	G	A	G	G	G	T	G	T	C	T	T	G	G	C	T	A	C	G	-	A	C	-	1	-	-	-	**1**
23	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	-	G	T	T	G	-	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	1	-	-	**1**
24	C	C	T	A	A	A	A	T	C	G	A	T	G	C	G	C	T	C	A	T	T	C	A	C	C	A	G	A	G	A	G	A	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	1	-	-	**1**
25	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	-	C	-	-	1	-	-	**1**
26	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	-	T	C	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	1	-	-	**1**
27	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	-	T	C	G	T	T	G	-	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	-	-	1	**1**
28	C	C	T	A	A	G	G	T	-	-	-	T	G	-	G	-	C	-	G	G	C	T	A	C	-	A	G	A	A	A	G	-	-	C	-	-	G	-	-	-	A	C	G	-	-	C	-	-	-	-	1	**1**
29	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	T	T	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	1	-	-	**1**
30	C	C	T	A	A	G	A	T	C	T	A	T	G	C	G	T	T	T	A	T	C	C	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	-	A	C	-	1	-	-	-	**1**
31	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	G	C	T	A	C	C	A	-	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	-	-	1	**1**
32	C	C	T	A	A	A	A	T	C	G	A	T	G	C	G	C	T	C	A	T	T	C	A	C	C	A	G	A	G	A	G	-	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	-	1	-	**1**
33	C	C	C	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	C	C	G	T	T	G	A	G	G	G	T	G	T	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	-	1	-	**1**
34	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	**1**
35	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	-	T	G	G	C	T	A	C	C	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	-	-	1	**1**
36	C	C	T	A	A	G	G	T	C	T	A	T	G	C	G	T	C	T	G	-	C	T	A	C	-	A	G	A	A	A	G	A	C	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	-	-	1	**1**
37	C	C	-	G	G	A	A	C	T	G	G	C	T	T	T	C	T	C	A	T	T	C	G	T	T	G	A	G	G	G	T	G	T	T	C	G	A	A	T	C	C	T	A	T	C	T	-	-	-	-	1	**1**
38	C	T	C	G	G	A	G	C	T	G	G	C	T	T	T	C	C	C	G	G	C	T	G	T	T	G	A	G	G	G	G	G	T	C	T	T	G	G	C	T	A	C	G	C	A	C	-	-	1	-	-	**1**
Number of genotypes																																															**11**	**46**	**54**	**25**	**79**	**215**

Table S7. Summary of multiple traits for 24 genotypes representing the RDM_I and _IV subgroups obtained by the GiA roots software

		ARW	NB	MNR	MeNR	NA	NP	SRL	NSA	NL	NV
ANOVA comparing RDM_I and RDM IV -summary-	p-value	n.s.	n.s.	0.00004 ***	0.00007 ***	0.000000003 ***	0.0000000003 ***	n.s.	0.000000001* **	0.00000000008 ***	0.00000004 ***
	mean RDM_I	0.056	1.90	13.205	7.219	17.121	700.65	321.51	65.439	368.724	1.309
	mean RDM_IV	0.055	2.02	10.133	5.422	10.163	416.39	347.43	38.112	212.668	0.747
	L.S.D.			2.069	1.246	3.256	125.37		12.562	66.242	0.287

Geno	RDM group	Rep	ARW	NB	MNR	MeNR	NA	NP	SRL	NSA	NL	NV
33	I	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
33	I	2	0.028	3.00	12	4	2.90	240	1310	10.9	124	0.09
33	I	3	0.054	1.57	11	7	18.45	797	330	69.4	411	1.24
33	I	4	0.058	1.82	20	11	28.73	1132	271	109.2	595	2.19
33	I	5	0.056	2.40	12	5	16.38	676	296	62.8	357	1.20
99	I	1	0.064	1.33	8	6	16.24	576	221	62.5	309	1.40
99	I	2	0.065	1.43	10	7	17.48	623	224	66.1	326	1.45
99	I	3	0.051	1.71	12	7	9.53	441	365	35.7	222	0.61
99	I	4	0.063	1.75	7	4	9.17	336	227	35.3	179	0.79
99	I	5	0.059	1.55	17	11	23.66	908	260	91.9	493	1.89
107	I	1	0.060	2.14	15	7	21.88	837	251	84.0	447	1.78
107	I	2	0.062	2.75	11	4	17.36	634	235	66.3	340	1.45
107	I	3	0.049	2.14	15	7	18.01	840	406	68.9	450	1.11
107	I	4	0.049	1.88	15	8	17.62	823	397	67.0	440	1.11
107	I	5	0.057	1.42	17	12	22.76	914	281	87.4	489	1.74
126	I	1	0.049	1.63	13	8	17.97	869	430	66.7	435	1.01
126	I	2	0.049	2.00	16	8	19.73	978	442	71.7	468	1.06
126	I	3	0.034	2.00	12	6	3.59	241	830	13.5	127	0.15
126	I	4	0.053	1.71	12	7	13.87	625	360	51.8	314	0.87
126	I	5	0.054	2.00	12	6	10.72	473	347	41.1	243	0.70
155	I	1	0.055	2.38	19	8	22.65	937	285	86.7	504	1.77
155	I	2	0.059	1.40	14	10	24.43	933	233	94.2	504	2.17
155	I	3	0.054	1.88	15	8	23.61	987	318	92.1	544	1.71
155	I	4	0.062	1.67	15	9	24.52	923	230	94.1	484	2.11
155	I	5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
241	IV	1	0.032	2.00	6	3	1.80	127	1004	6.7	65	0.07
241	IV	2	0.053	1.00	1	1	0.86	37	441	3.0	18	0.04

241	IV	3	0.062	2.33	7	3	9.70	348	228	36.7	188	0.82
241	IV	4	0.066	1.38	11	8	16.65	585	227	61.9	298	1.31
241	IV	5	0.064	1.33	4	3	4.51	160	258	16.2	80	0.31
269	IV	1	0.046	1.40	7	5	5.94	311	481	21.4	148	0.31
269	IV	2	0.058	1.80	9	5	7.36	293	265	28.1	155	0.59
269	IV	3	0.056	1.40	7	5	7.73	322	292	28.9	163	0.56
269	IV	4	0.058	1.67	5	3	4.10	170	282	15.1	83	0.30
269	IV	5	0.053	4.00	8	2	5.70	238	302	22.0	132	0.44
287	IV	1	0.065	2.00	10	5	10.85	394	194	42.8	209	1.08
287	IV	2	0.069	1.50	9	6	9.63	327	160	38.8	179	1.12
287	IV	3	0.052	2.11	19	9	13.19	589	321	50.3	308	0.96
287	IV	4	0.057	1.75	14	8	12.08	489	276	45.5	252	0.91
287	IV	5	0.059	2.33	7	3	6.64	259	260	26.3	142	0.55
292	IV	1	0.061	2.13	17	8	13.44	508	258	49.5	257	1.00
292	IV	2	0.050	4.50	9	2	9.20	433	404	34.2	217	0.54
292	IV	3	0.057	1.75	14	8	12.94	515	306	48.9	274	0.90
292	IV	4	0.057	1.50	9	6	9.21	386	296	33.9	189	0.64
292	IV	5	0.058	2.33	14	6	12.31	479	280	46.6	254	0.91
311	I	1	0.056	2.13	17	8	21.57	880	275	84.4	477	1.74
311	I	2	0.054	1.57	11	7	14.61	639	328	54.6	321	0.98
311	I	3	0.061	1.60	16	10	24.04	892	245	93.8	487	1.99
311	I	4	0.064	2.00	8	4	13.77	487	207	53.2	263	1.27
311	I	5	0.057	1.50	9	6	13.98	557	281	54.3	303	1.08
325	IV	1	0.057	1.56	14	9	21.68	902	302	79.5	444	1.47
325	IV	2	0.057	1.60	8	5	9.48	402	280	35.2	197	0.70
325	IV	3	0.055	1.70	17	10	22.15	948	330	81.9	476	1.44
325	IV	4	0.054	2.50	15	6	15.35	689	326	56.9	336	1.03
325	IV	5	0.065	1.63	13	8	17.48	624	208	65.7	323	1.55
327	IV	1	0.050	2.80	14	5	12.43	609	412	45.3	288	0.70
327	IV	2	0.054	2.00	12	6	9.53	421	312	35.4	207	0.66
327	IV	3	0.056	1.80	9	5	9.03	375	309	33.7	191	0.62
327	IV	4	0.063	1.33	4	3	3.10	113	255	11.5	58	0.23
327	IV	5	0.060	2.00	8	4	6.76	252	255	26.4	141	0.55
339	IV	1	0.057	2.20	11	5	12.10	484	267	46.9	261	0.97
339	IV	2	0.054	2.25	9	4	11.37	509	333	42.2	250	0.75
339	IV	3	0.056	1.75	14	8	18.15	771	321	67.2	385	1.20
339	IV	4	0.057	2.00	8	4	7.08	290	312	26.2	147	0.47
339	IV	5	0.040	1.60	8	5	2.14	123	604	8.0	64	0.11
378	I	1	0.057	3.25	13	4	13.90	567	266	53.0	296	1.11
378	I	2	0.054	2.00	12	6	15.45	667	307	59.6	353	1.15
378	I	3	0.053	2.71	19	7	21.29	935	329	80.0	479	1.46

378	I	4	0.053	1.67	15	9	18.77	837	337	70.0	418	1.24
378	I	5	0.061	2.00	14	7	19.85	772	209	78.1	405	1.93
382	I	1	0.059	2.00	12	6	17.18	674	271	65.9	355	1.31
382	I	2	0.060	3.00	15	5	20.77	764	237	80.9	428	1.81
382	I	3	0.055	2.14	15	7	21.40	916	299	79.5	457	1.53
382	I	4	0.058	1.63	13	8	17.74	701	275	67.9	374	1.36
382	I	5	0.056	1.80	18	10	25.72	1017	279	101.8	579	2.07
384	I	1	0.066	1.38	11	8	14.18	475	190	55.5	268	1.41
384	I	2	0.046	1.60	16	10	12.15	612	435	47.0	324	0.74
384	I	3	0.056	1.56	14	9	19.75	814	298	74.7	424	1.43
384	I	4	0.058	1.50	12	8	17.48	703	282	65.5	361	1.28
384	I	5	0.056	1.38	18	13	22.69	929	281	85.4	487	1.73
388	I	1	0.049	2.50	15	6	12.40	611	433	44.9	295	0.68
388	I	2	0.061	1.38	11	8	15.79	596	267	58.6	307	1.15
388	I	3	0.053	2.00	10	5	10.89	496	349	40.0	242	0.69
388	I	4	0.054	2.00	10	5	11.87	509	331	44.6	265	0.80
388	I	5	0.055	3.25	13	4	13.82	559	290	53.2	309	1.07
392	I	1	0.064	2.75	11	4	12.26	452	231	44.9	223	0.96
392	I	2	0.060	2.20	11	5	14.96	564	255	56.9	303	1.19
392	I	3	0.036	1.43	10	7	4.40	291	764	16.3	145	0.19
392	I	4	0.056	1.43	10	7	18.14	756	279	71.7	409	1.47
392	I	5	0.056	1.50	9	6	15.38	639	289	58.8	335	1.16
393	I	1	0.055	1.86	13	7	14.67	645	332	53.5	310	0.93
393	I	2	0.047	1.60	8	5	9.97	510	479	36.2	244	0.51
393	I	3	0.070	1.75	7	4	10.69	350	178	41.9	191	1.07
393	I	4	0.057	2.00	8	4	6.27	256	311	23.0	128	0.41
393	I	5	0.056	1.63	13	8	17.18	723	293	64.9	369	1.26
398	I	1	0.059	2.22	20	9	25.96	1022	254	99.4	536	2.11
398	I	2	0.057	2.14	15	7	23.48	952	287	90.2	507	1.77
398	I	3	0.061	1.63	13	8	19.44	750	244	73.8	383	1.57
398	I	4	0.064	2.14	15	7	17.56	586	199	70.0	346	1.74
398	I	5	0.070	1.40	14	10	21.24	673	180	82.8	379	2.10
411	I	1	0.057	2.60	13	5	14.85	617	306	55.8	311	1.01
411	I	2	0.060	1.75	14	8	24.33	930	250	94.4	503	2.02
411	I	3	0.054	1.50	12	8	17.08	740	322	64.7	379	1.18
411	I	4	0.058	1.67	15	9	21.59	844	275	83.2	454	1.65
411	I	5	0.062	1.83	11	6	18.69	715	233	71.1	363	1.56
419	I	1	0.040	1.57	11	7	6.11	343	566	23.8	190	0.34
419	I	2	0.055	2.22	20	9	22.44	957	302	85.1	493	1.63
419	I	3	0.056	1.50	12	8	16.31	688	292	62.0	354	1.21
419	I	4	0.055	1.91	21	11	26.70	1094	293	103.3	595	2.03

419	I	5	0.061	1.38	11	8	17.77	703	205	73.9	388	1.90
582	IV	1	0.059	1.67	5	3	5.68	221	293	20.8	111	0.38
582	IV	2	0.033	1.50	9	6	3.57	251	894	13.2	125	0.14
582	IV	3	0.053	1.67	10	6	6.51	279	319	24.6	148	0.46
582	IV	4	0.055	2.00	10	5	9.67	420	327	35.6	206	0.63
582	IV	5	0.054	2.29	16	7	11.74	498	331	44.0	262	0.79
586	IV	1	0.061	1.50	9	6	15.23	578	253	56.2	293	1.16
586	IV	2	0.060	1.60	8	5	7.46	285	265	28.1	149	0.56
586	IV	3	0.060	1.60	16	10	16.00	598	248	61.4	323	1.30
586	IV	4	0.056	2.75	11	4	12.49	524	312	47.1	268	0.86
586	IV	5	0.068	1.83	11	6	17.40	600	201	65.2	305	1.52

ARW = Average root width, NB = Network bushiness, MNR = Maximum number of roots, MeNR = Median number of roots, NA = Network area, NP = Network perimeter, SRL = Specific root length, NSA = Network surface area, NLD = Network length distribution, NV = Network volume

Table S8. SNP annotations and orthologues on rice and sorghum

Rice chromosome	Rice annotation	Rice feature	Rice start	Rice end	Rice IRGSP gene ID	Wheat ortholog (ENSEMBL30)	Wheat start (if 5B)	Wheat end (if 5B)	Rice MSU ID	Sorghum orthlog	Sorghum chromosome	Sorghum start	Sorghum end	Rice MSU functional annotation
	irgs	gen	335300	335344	OS03G0803	Traes_5BL_8E038ABF	2469315	2469420	LOC_Os03g58	Sb04g0332		631953		
3	p	e	88	23	700	F.1	63	39	910	93	4	65	63195870	expressed protein
	irgs	gen	335391	335445	OS03G0803	Traes_5BL_8E038ABF	2469315	2469420	LOC_Os03g58	Sb04g0332		631953		
3	p	e	06	90	800	F.1	63	39	910	93	4	65	63195870	expressed protein
	irgs	gen	335465	335510	OS03G0803	Traes_5BL_2B137911F	2449004	2449034	LOC_Os03g58	Sb01g0046		370273		
3	p	e	74	64	900	.1	10	34	920	30	1	9	3707122	galactosyltransferase family protein, putative, expressed
	irgs	gen	335532	335542	OS03G0804									
3	p	e	83	30	000				None					
	irgs	gen	335536	335563	OS03G0804	Traes_5BL_E40C680C	2476644	2476657	LOC_Os03g58	Sb01g0046		369538		
3	p	e	33	54	100	0.1	60	59	930	20	1	4	3697749	expressed protein
	irgs	gen	335637	335666	OS03G0804	Traes_5BL_C73A7674	2492716	2492738	LOC_Os03g58	Sb01g0046		368736		LTPL83 - Protease inhibitor/seed storage/LTP family protein precursor,
3	p	e	02	13	200	6.1	36	28	940	10	1	6	3689943	putative, expressed
	irgs	gen	335705	335761	OS03G0804	Traes_2BS_6402A1C5			LOC_Os03g58	Sb06g0040		895641		
3	p	e	53	13	300	7.1			960	40	6	6	8962237	DHHC zinc finger domain containing protein, expressed
	irgs	gen	335794	335851	OS03G0804	Traes_5BL_8B6FE75D	2461029	2461070	LOC_Os03g58	Sb01g0046		368006		
3	p	e	93	33	400	7.1	99	32	970	00	1	8	3685443	expressed protein
	irgs	gen	335864	335873	OS03G0804	Traes_5BL_B92355534			LOC_Os03g58	Sb01g0045		367838		
3	p	e	65	20	500	1.1			980	90	1	0	3679397	Cupin domain containing protein, expressed
	irgs	gen	335885	335894	OS03G0804				LOC_Os03g58					
3	p	e	52	11	600				990					cupin domain containing protein, expressed
	irgs	gen	335911	335918	OS03G0804	Traes_5BL_B92355534			LOC_Os03g59					
3	p	e	95	72	700	1.1			010					Cupin domain containing protein, expressed
	irgs	gen	335927	335980	OS03G0804	Traes_4AL_BF03B5E0			LOC_Os03g59	Sb05g0224		543739		
3	p	e	86	72	800	9.2			020	70	5	81	54378736	T-complex protein, putative, expressed
	irgs	gen	335990	336009	OS03G0804	Traes_4AL_EC565F87			LOC_Os03g59	Sb01g0045		366536		
3	p	e	55	84	900	F.1			030	70	1	1	3666836	UDP-rhamnose rhamnosyltransferase, putative, expressed
	irgs	gen	336055	336124	OS03G0805	Traes_5BL_CC637257	2484435	2484476	LOC_Os03g59	Sb01g0045		363124		
3	p	e	47	23	100	C.1	93	20	040	50	1	2	3637745	squalene synthetase, putative, expressed
	irgs	gen	336125	336177	OS03G0805	Traes_5BL_D8099BFB			LOC_Os03g59	Sb01g0045		362587		
3	p	e	11	79	200	6.2			050	40	1	6	3630837	DEAD-box ATP-dependent RNA helicase, putative, expressed
	irgs	gen	336219	336265	OS03G0805	Traes_5BL_50C5F7E9	2497435	2497480	LOC_Os03g59	Sb01g0045		361889		
3	p	e	70	03	300	C.1	46	97	060	30	1	4	3623287	OsPP2Ac-2 - Phosphatase 2A isoform 2 belonging to family 2, expressed
	irgs	gen	336249	336259	OS03G0805									
3	p	e	20	92	350				None					
	irgs	gen	336273	336302	OS03G0805	Traes_5BL_18D3F8F3	2470568	2470599	LOC_Os03g59	Sb01g0045		361481		
3	p	e	74	85	400	8.1	75	34	070	20	1	9	3618019	phosphatase, putative, expressed
	irgs	gen	336319	336368	OS03G0805				LOC_Os03g59					
3	p	e	31	72	500				080					AMP-binding enzyme, putative, expressed
	irgs	gen	336386	336411	OS03G0805	Traes_5BL_36992DC0			LOC_Os03g59	Sb01g0044		355054		
3	p	e	50	23	600	F.1			090	40	1	6	3553033	retrotransposon protein, putative, unclassified, expressed
	irgs	gen	336451	336475	OS03G0805	Traes_7AS_6200CA75			LOC_Os03g59	Sb01g0044		355054		
3	p	e	07	19	700	A.1			100	40	1	6	3553033	pheophorbide a oxygenase, chloroplast precursor, putative, expressed
	irgs	gen	336451	336475	OS03G0805									
3	p	e	07	19	733				None					
	irgs	gen	336493	336537	OS03G0805									
3	p	e	42	55	766				None					
	irgs	gen	336508	336527	OS03G0805				LOC_Os03g59					
3	p	e	72	10	800				110					pheophorbide a oxygenase, chloroplast precursor, putative, expressed
	irgs	gen	336595	336619	OS03G0805									
3	p	e	65	54	850				None					

	irgs	gen	336595	336620	OS03G0805	Traes_3B_D50AE46FF.			LOC_Os03g59	Sb01g0044		355054			
3	p	e	67	28	900	1			120	40	1	6	3553033	pheophorbide a oxygenase, chloroplast precursor, putative, expressed	
	irgs	gen	336808	336874	OS03G0806										
3	p	e	75	66	100				None						
	irgs	gen	336808	336877	OS03G0806	Traes_5AL_5936BB11			LOC_Os03g59	Sb01g0044		355054			
3	p	e	75	37	300	0.1			146	40	1	6	3553033	expressed protein	
	irgs	gen	336880	336923	OS03G0806				LOC_Os03g59						
3	p	e	13	14	400				160					elongation factor P, putative, expressed	
	irgs	gen	336931	336959	OS03G0806	Traes_5BL_377B23238	2495323	2495346	LOC_Os03g59	Sb01g0044		353846			
3	p	e	71	57	500	.1	66	30	170	20	1	9	3541785	OsAPRL5 adenosine 5'-phosphosulfate reductase-like OsAPRL5, expressed	
	irgs	gen	336971	336984	OS03G0806	Traes_5BL_53A573FB	2495300	2495312	LOC_Os03g59	Sb01g0044		353546			
3	p	e	76	42	600	B.2	42	80	180	10	1	2	3536504	expressed protein	
	irgs	gen	337114	337135	OS03G0806	Traes_5BL_EE791D41			LOC_Os03g59	Sb01g0044		352184			
3	p	e	73	42	700	9.2			200	00	1	2	3522898	expressed protein	
	irgs	gen	337169	337177	OS03G0806	Traes_5BL_4B68B4F4	2480791	2480802	LOC_Os03g59						
3	p	e	07	25	800	8.2	09	95	210					expressed protein	
	irgs	gen	337185	337205	OS03G0806	Traes_5BL_B098BF80			LOC_Os03g59	Sb01g0043		351589			
3	p	e	60	90	900	7.1			220	90	1	4	3518747	cytochrome b-c1 complex subunit 7, putative, expressed	
	irgs	gen	337220	337248	OS03G0807	Traes_5BL_174D06810	2477008	2477035	LOC_Os03g59	Sb01g0043		351495			
3	p	e	11	50	000	.2	10	44	225	80	1	2	3515302	expressed protein	
	irgs	gen	337315	337349	OS03G0807	Traes_5BL_025C74BD	2450661	2450697	LOC_Os03g59	Sb01g0043		348123			
3	p	e	66	58	100	B.1	84	51	240	60	1	4	3485337	carboxyl-terminal peptidase, putative, expressed	
	irgs	gen	337323	337349	OS03G0807										
3	p	e	91	27	150				None						
	irgs	gen	337374	337394	OS03G0807	Traes_5BL_3A6B74A8	2459593	2459617	LOC_Os03g59	Sb01g0043		347076			
3	p	e	58	97	200	0.2	22	89	250	40	1	9	3473239	expressed protein	
	irgs	gen	337440	337494	OS03G0807	Traes_7AS_404BDBB4			LOC_Os03g59	Sb07g0013		136010			
3	p	e	48	91	400	8.1			264	80	7	3	1363190	calreticulin family protein, expressed	
	irgs	gen	337527	337560	OS03G0807	Traes_5BL_BFF0FD47	2448479	2448497	LOC_Os03g59	Sb01g0043		346082			
3	p	e	99	68	500	D.1	19	62	280	20	1	3	3463382	plastocyanin-like domain containing protein, putative, expressed	
	irgs	gen	337561	337586	OS03G0807				LOC_Os03g59	Sb01g0043		345939			
3	p	e	72	45	600				290	10	1	7	3460374	C-methyltransferase, putative, expressed	
	irgs	gen	337657	337677	OS03G0807	Traes_5BL_7873A9BC	2457890	2457912	LOC_Os03g59	Sb01g0042		341968			
3	p	e	12	00	700	0.2	00	22	300	70	1	9	3422032	expressed protein	
	irgs	gen	337690	337708	OS03G0807	Traes_3B_E3B333A45.			LOC_Os03g59	Sb10g0201		432613			
3	p	e	18	93	800	1			310	50	10	20	43263008	ribosomal protein, putative, expressed	
	irgs	gen	337717	337735	OS03G0807	Traes_5BL_D203F3948	2448570	2448589	LOC_Os03g59	Sb01g0042		340366			
3	p	e	56	81	900	.1	46	08	320	40	1	9	3404860	expressed protein	
	irgs	gen	337774	337803	OS03G0808	Traes_5BL_98B592A5			LOC_Os03g59	Sb01g0041		336331			
3	p	e	71	50	000	C.1			330	90	1	8	3364904	polygalacturonase, putative, expressed	
	irgs	gen	337800	337862	OS03G0808	Traes_5BL_CFCBFDA	2470701	2470752	LOC_Os03g59	Sb01g0042		338594			
3	p	e	15	59	100	99.2	19	16	340	10	1	7	3391819	CESA2 - cellulose synthase, expressed	
	irgs	gen	337806	337813	OS03G0808				LOC_Os03g59						
3	p	e	04	85	150				330					polygalacturonase, putative, expressed	
	irgs	gen	337879	337894	OS03G0808	Traes_5BL_529801541.	2180577	2180588	LOC_Os03g59	Sb01g0083		720840			
3	p	e	09	17	200	1	67	34	350	70	1	0	7209740	anthocyanin 3-O-beta-glucosyltransferase, putative, expressed	
	irgs	gen	337910	337927	OS03G0808	Traes_5DL_753A1DC			LOC_Os03g59	Sb01g0042		337936			
3	p	e	08	33	300	A0.1			360	00	1	7	3381567	remorin, putative, expressed	
	irgs	gen	337912	337926	OS03G0808										
3	p	e	88	25	350				None						
	irgs	gen	337967	337978	OS03G0808	Traes_5BL_481EF6F2			LOC_Os03g59	Sb01g0041		336077			
3	p	e	89	66	400	A.1			370	80	1	5	3362611	ubiquitin family protein, putative, expressed	
	irgs	gen	337989	337996	OS03G0808	Traes_5BL_B7B05C9A			LOC_Os03g59					LTPL28 - Protease inhibitor/seed storage/LTP family protein precursor,	
3	p	e	40	87	500	F1.1			380					expressed	
	irgs	gen	338058	338100	OS03G0808	Traes_5BL_35A6B438	2505993	2506027	LOC_Os03g59	Sb01g0041		334443		CAMK_CAMK_like.24 - CAMK includes calcium/calmodulin depedent	
3	p	e	47	79	600	7.2	35	91	390	50	1	2	3348438	protein kinases, expressed	
	irgs	gen	338062	338097	OS03G0808										
3	p	e	57	75	750				None						
	irgs	gen	338295	338341	OS03G0808	Traes_3AS_7F8EB9A5			LOC_Os03g59	Sb01g0041		332732			
3	p	e	10	46	900	7.1			430	30	1	1	3332332	uncharacterized glycosyltransferase, putative, expressed	
	irgs	gen	338297	338340	OS03G0808										
3	p	e	09	13	950				None						
	irgs	gen	338346	338354	OS03G0809	Traes_3AL_7899DC65			LOC_Os03g59	Sb01g0041		332663			
3	p	e	38	99	000	1.1			440	20	1	8	3327186	dirigent, putative, expressed	
	irgs	gen	338379	338416	OS03G0809	Traes_5BL_6032BE9D	2499198	2499237	LOC_Os03g59	Sb01g0041		331636			
3	p	e	42	46	100	2.2	63	25	450	00	1	6	3318406	transporter-related, putative, expressed	
	irgs	gen	338442	338455	OS03G0809	Traes_3B_CB97D8DB			LOC_Os03g59	Sb01g0040		330449			
3	p	e	23	80	200	3.1			460	80	1	0	3307002	transcription factor, putative, expressed	

	irgs	gen	338457	338498	OS03G0809	Traes_5BL_B13010DB	2506704	2506735	LOC_Os03g59	Sb01g0040		328195		
3	p	e	82	47	300	D.1	28	04	470	30	1	9	3287477	stage II sporulation protein E, putative, expressed
	irgs	gen	338503	338552	OS03G0809				LOC_Os03g59	Sb06g0098		277110		
3	p	e	63	73	400				480	20	6	28	27711941	expressed protein
	irgs	gen	338868	338918	OS03G0809	Traes_5BL_04CF9526F	2453756	2453803	LOC_Os03g59					
3	p	e	20	84	700	.1	58	06	530					protein Kinase-associated protein phosphatase, putative, expressed
	irgs	gen	338919	338932	OS03G0809	Traes_5BL_DBB9A854	2478437	2478460	LOC_Os03g59	Sb01g0040		327056		
3	p	e	72	70	800	6.2	74	10	540	16	1	7	3272988	zinc finger, C3HC4 type domain containing protein, expressed
	irgs	gen	338948	338996	OS03G0809	Traes_5BL_091E1C819	2453230	2453256	LOC_Os03g59	Sb01g0040		326160		
3	p	e	09	21	900	.1	53	57	550	10	1	2	3265980	RNA recognition motif containing protein, putative, expressed
	irgs	gen	339058	339074	OS03G0810	Traes_5BL_CC443DA6	2489099	2489115	LOC_Os03g59					
3	p	e	26	96	100	C.1	72	62	570					IPP transferase, putative, expressed
	irgs	gen	339168	339212	OS03G0810				LOC_Os03g59	Sb01g0039		318655		
3	p	e	21	05	300				580	40	1	5	3190220	hydrolase, NUDIX family, domain containing protein, expressed
	irgs	gen	339264	339271	OS03G0810									
3	p	e	82	10	500				None					
	irgs	gen	339276	339338	OS03G0810	Traes_5BL_808FF9377	2560624	2560686	LOC_Os03g59	Sb01g0039		314399		
3	p	e	12	44	600	.1	79	69	590	00	1	6	3151010	ATP/GTP/Ca++ binding protein, putative, expressed
	irgs	gen	339361	339376	OS03G0810	Traes_3B_1F02C3FEE.			LOC_Os03g59					
3	p	e	55	15	700	2			600					mitochondrial Rho GTPase 1, putative, expressed
	irgs	gen	339389	339413	OS03G0810	Traes_5BL_14E1BB21			LOC_Os03g59	Sb01g0038		313148		oxidoreductase, short chain dehydrogenase/reductase family protein, putative,
3	p	e	98	23	800	9.1			610	80	1	4	3133196	expressed
	irgs	gen	339402	339410	OS03G0810									
3	p	e	65	21	850				None					
	irgs	gen	339416	339461	OS03G0810	Traes_5BL_FF1CDA02	2547983	2548023	LOC_Os03g59	Sb01g0038		312765		
3	p	e	95	47	900	0.1	75	57	620	70	1	2	3131443	phospholipase, patatin family, putative, expressed
	irgs	gen	339552	339619	OS03G0811	Traes_5BL_1DE088A3	2524289	2524319	LOC_Os03g59	Sb01g0038		311477		
3	p	e	93	62	100	0.2	20	00	640	60	1	4	3124068	magnesium-chelatase subunit chlD, chloroplast precursor, putative, expressed
	irgs	gen	339623	339680	OS03G0811	Traes_5BL_47EB5C6B	2522314	2522374	LOC_Os03g59	Sb01g0038		310963		
3	p	e	73	64	200	2.1	79	70	650	50	1	7	3113506	BRCA1 C Terminus domain containing protein, expressed
	irgs	gen	339691	339715	OS03G0811	Traes_4DL_4F2B74143			LOC_Os03g59	Sb01g0038		309787		
3	p	e	67	19	300	.1			660	40	1	6	3100100	clathrin adaptor complex small chain domain containing protein, expressed
	irgs	gen	339737	339755	OS03G0811	Traes_5BL_CBFAD4B	2548302	2548324	LOC_Os03g59	Sb01g0038		309526		
3	p	e	60	25	400	59.1	63	90	670	30	1	3	3096583	basic helix-loop-helix, putative, expressed
	irgs	gen	339882	339919	OS03G0811	Traes_5BL_2377038E	2553897	2553945	LOC_Os03g59	Sb01g0038		307109		
3	p	e	16	06	500	D.1	24	42	680	20	1	4	3074942	PAPA-1-like conserved region family protein, expressed
	irgs	gen	339943	339953	OS03G0811	Traes_5BL_06084CDA			LOC_Os03g59	Sb01g0038		305562		
3	p	e	25	05	550	61.1			690	10	1	3	3056567	DUF617 domain containing protein, expressed
	irgs	gen	339956	339986	OS03G0811	Traes_5BL_8054CBDB	2554865	2554899	LOC_Os03g59	Sb01g0038		305179		
3	p	e	33	31	600	6.1	51	80	700	00	1	1	3055291	peptidyl-prolyl cis-trans isomerase, putative, expressed
	irgs	gen	339988	339994	OS03G0811	Traes_5BL_DB83FBE9	2554853	2554862	LOC_Os03g59	Sb01g0037		305011		
3	p	e	08	67	700	11.1	28	85	710	90	1	9	3051515	RNA recognition motif containing protein, putative, expressed
	irgs	gen	340010	340023	OS03G0811	Traes_5BL_A89BFEF6	2554817	2554838	LOC_Os03g59	Sb01g0037		304803		
3	p	e	75	88	800	C.2	30	61	720	80	1	8	3049672	ribosomal protein L36 containing protein, expressed
	irgs	gen	340069	340096	OS03G0811	Traes_1DL_34A359364			LOC_Os03g59	Sb01g0037		303445		
3	p	e	99	31	900	.1			740	70	1	3	3037078	ADP-ribosylation factor, putative, expressed
	irgs	gen	340109	340156	OS03G0812	Traes_5BL_A09FB407	2538079	2538116		Sb01g0037		301947		
3	p	e	18	42	000	5.1	70	45	None	50	1	0	3031368	
	irgs	gen	340314	340340	OS03G0812	Traes_5BL_9F5869751.	2532128	2532156	LOC_Os03g59	Sb01g0037		301086		
3	p	e	52	30	200	1	77	32	760	40	1	1	3013787	RING finger protein 126, putative, expressed
	irgs	gen	340343	340351	OS03G0812									
3	p	e	93	01	300				None					
	irgs	gen	340362	340370	OS03G0812	Traes_1DL_C3C407FC			LOC_Os03g59	Sb01g0037		300674		
3	p	e	46	92	400	1.1			770	30	1	2	3007425	EF hand family protein, putative, expressed
	irgs	gen	340504	340559	OS03G0812	Traes_1DL_696E3F6A			LOC_Os03g59	Sb01g0037		300674		
3	p	e	46	50	800	2.1			790	30	1	2	3007425	EF hand family protein, putative, expressed
	irgs	gen	340716	340723	OS03G0813	Traes_4AL_491C00040			LOC_Os03g59	Sb01g0017		153991		
3	p	e	01	69	200	.1			840	70	1	5	1542434	expressed protein
	irgs	gen	340889	340897	OS03G0813	Traes_6BS_257A12B3			LOC_Os03g59					
3	p	e	86	87	700	4.1			880					expressed protein
	irgs	gen	341080	341085	OS03G0814				LOC_Os03g59					
3	p	e	14	87	200				940					expressed protein
	irgs	gen	341235	341243	OS03G0814				LOC_Os03g59					
3	p	e	36	20	500				990					glucosyl transferase, putative, expressed
	irgs	gen	341239	341244	OS03G0814									
3	p	e	76	43	550				None					
	irgs	gen	341257	341282	OS03G0814	Traes_4AL_150EF1F98			LOC_Os03g60	Sb01g0017		153991		
3	p	e	40	14	600	.1			000	70	1	5	1542434	SPFH domain/Band 7 family protein, expressed

	irgs	gen	341447	341452	OS03G0814				LOC_Os03g60	Sb07g0025		281830			
3	p	e	37	71	800				030	80	7	8	2819330	transposon protein, putative, CACTA, En/Spm sub-class, expressed	
	irgs	gen	341661	341675	OS03G0815	Traes_5BL_4497A137	2501439	2501460	LOC_Os03g60	Sb01g0037		298184			
3	p	e	00	21	100	C.1	56	03	080	10	1	8	2983430	NAC domain-containing protein 67, putative, expressed	
	irgs	gen	341678	341733	OS03G0815	Traes_5BL_8C56FDC6	2559570	2559626	LOC_Os03g60	Sb01g0037		297579			
3	p	e	24	28	200	4.1	45	80	090	00	1	5	2981369	methylenetetrahydrofolate reductase, putative, expressed	
	irgs	gen	341746	341750	OS03G0815										
3	p	e	82	47	332				None						
	irgs	gen	341747	341766	OS03G0815	Traes_5BL_AA591194	2548133	2548157	LOC_Os03g60	Sb01g0036		297188			
3	p	e	79	67	400	9.1	82	44	100	90	1	0	2974178	50S ribosomal protein L17, putative, expressed	
	irgs	gen	341864	341907	OS03G0815	Traes_5BL_EC8906F8	2525392	2525429	LOC_Os03g60	Sb01g0036		296085			
3	p	e	37	37	700	9.1	42	65	110	80	1	5	2964672	KH domain containing protein, putative, expressed	
	irgs	gen	341925	341936	OS03G0815				LOC_Os03g60	Sb01g0036		295759			
3	p	e	60	50	800				120	70	1	0	2958737	AP2 domain containing protein, expressed	
	irgs	gen	341952	341979	OS03G0815	Traes_5BL_5897DA40	2538363	2538412	LOC_Os03g60	Sb01g0036		295044			
3	p	e	26	90	900	4.1	09	26	130	60	1	9	2953477	transcription elongation factor protein, putative, expressed	
	irgs	gen	341988	342018	OS03G0816				LOC_Os03g60	Sb01g0036		294642			
3	p	e	79	89	000				140	50	1	7	2949876	U-box domain-containing protein, putative, expressed	
	irgs	gen	342024	342062	OS03G0816	Traes_5BL_94E74ECF	2529063	2529104	LOC_Os03g60	Sb01g0036		294139			
3	p	e	64	71	100	E.2	05	78	150	40	1	9	2945685	protein kinase domain containing protein, expressed	
	irgs	gen	342071	342076	OS03G0816										
3	p	e	76	68	150				None						
	irgs	gen	342154	342181	OS03G0816	Traes_5BL_EC6277E0	2524923	2524945		Sb01g0036		293252			
3	p	e	10	00	200	E.2	40	70	None	30	1	6	2935029		
	irgs	gen	342199	342226	OS03G0816	Traes_1DL_3259FC4A			LOC_Os03g60						
3	p	e	47	83	300	0.2			170					expressed protein	
	irgs	gen	342275	342321	OS03G0816	Traes_5BL_7575A8789	2524849	2524907	LOC_Os03g60	Sb01g0036		292698			
3	p	e	28	33	400	.2	68	48	180	20	1	9	2931489	GTPase of unknown function domain containing protein, putative, expressed	
	irgs	gen	342344	342384	OS03G0816	Traes_5BL_0A9B605F	2525789	2525824	LOC_Os03g60	Sb01g0036		291204			
3	p	e	46	06	500	B.1	03	23	190	00	1	9	2914460	oxidoreductase, 2OG-Fe oxygenase family protein, putative, expressed	
	irgs	gen	342384	342413	OS03G0816	Traes_3B_EBAA32F3			LOC_Os03g60	Sb01g0131		121664			
3	p	e	74	83	600	B.2			200	30	1	54	12168784	pentatricopeptide, putative, expressed	
	irgs	gen	342421	342434	OS03G0816	Traes_5BL_F768416E	2550399	2550409	LOC_Os03g60	Sb01g0035		291071			
3	p	e	64	19	700	A.1	40	09	210	90	1	6	2911477	DUF567 domain containing protein, putative, expressed	
	irgs	gen	342450	342462	OS03G0816	Traes_5BL_57C4F661	2555900	2555918	LOC_Os03g60	Sb01g0035		290816			
3	p	e	96	12	800	E.1	09	60	220	80	1	4	2909008	DUF567 domain containing protein, putative, expressed	
	irgs	gen	342578	342635	OS03G0816	Traes_5AL_6AAF021C			LOC_Os03g60	Sb01g0035		289444			
3	p	e	58	71	900	F.2			240	70	1	6	2904075	SCAR-like protein 1, putative, expressed	
	irgs	gen	342602	342663	OS03G0816										
3	p	e	79	63	950				None						
	irgs	gen	342649	342668	OS03G0817	Traes_5AL_6AAF021C			LOC_Os03g60						
3	p	e	73	62	000	F.2			240					SCAR-like protein 1, putative, expressed	
	irgs	gen	342674	342687	OS03G0817	Traes_5BL_5532FFF54	2559255	2559263	LOC_Os03g60	Sb01g0035		289310			
3	p	e	83	11	100	.1	08	62	250	60	1	9	2894418	membrane associated DUF588 domain containing protein, putative, expressed	
	irgs	gen	342713	342730	OS03G0817	Traes_5BL_672642904.	2542456	2542473	LOC_Os03g60	Sb01g0035		287696			
3	p	e	61	67	200	1	19	09	260	40	1	9	2878880	ANT1, putative, expressed	
	irgs	gen	342898	342927	OS03G0817				LOC_Os03g60	Sb01g0035		285299			
3	p	e	59	82	500				300	30	1	0	2854694	expressed protein	
	irgs	gen	343051	343087	OS03G0817										
3	p	e	36	44	700				None						
	irgs	gen	343112	343147	OS03G0817	Traes_5BL_00D162E9	2534605	2534633	LOC_Os03g60	Sb01g0035		284540			
3	p	e	82	97	800	D.2	93	52	340	20	1	4	2848873	leaf senescence related protein, putative, expressed	
	irgs	gen	343205	343253	OS03G0817	Traes_5BL_E0C2C7D7	2523081	2523146	LOC_Os03g60	Sb01g0035		283387			
3	p	e	18	81	900	2.2	95	69	350	10	1	6	2837854	leaf senescence related protein, putative, expressed	
	irgs	gen	343271	343318	OS03G0818	Traes_5BL_18A9BA43	2508290	2508336	LOC_Os03g60	Sb01g0035		282454			
3	p	e	89	66	000	7.2	57	71	360	00	1	5	2828938	protein kinase PKN/PRK1, effector, putative, expressed	
	irgs	gen	343332	343334	OS03G0818										
3	p	e	45	81	050				None						
	irgs	gen	343347	343401	OS03G0818	Traes_5BL_347A9D93	2557611	2557667	LOC_Os03g60	Sb01g0034		281650			
3	p	e	56	33	100	7.1	58	77	370	90	1	0	2821730	histidine acid phosphatase, putative, expressed	
	irgs	gen	343402	343432	OS03G0818	Traes_5BL_D5FFA0E4	2553752	2553779	LOC_Os03g60	Sb01g0034		281370			
3	p	e	87	11	200	D.1	43	10	380	80	1	5	2815878	cinnamoyl CoA reductase, putative, expressed	
	irgs	gen	343431	343467	OS03G0818	Traes_5BL_077E68865			LOC_Os03g60	Sb01g0034		274161			
3	p	e	47	35	300	.1			390	20	1	8	2745295	PHD finger protein, putative, expressed	
	irgs	gen	343471	343494	OS03G0818	Traes_5BL_5EB1D581	2523390	2523427	LOC_Os03g60	Sb01g0034		273794			
3	p	e	78	15	400	3.1	64	37	400	10	1	0	2739951	40S ribosomal protein S23, putative, expressed	
	irgs	gen	343506	343529	OS03G0818										
3	p	e	04	15	700				None						

	irgs	gen	343581	343623	OS03G0818	Traes_5BL_5117B3E2	2535330	2535360	LOC_Os03g60	Sb01g0034		273112		
3	p	e	93	34	800	A.2	24	34	430	00	1	2	2735193	AP2 domain containing protein, expressed
	irgs	gen	343781	343823	OS03G0819	Traes_5BL_474FF094C	2540435	2540465	LOC_Os03g60	Sb01g0033		270204		
3	p	e	57	54	100	.1	19	74	460	80	1	2	2709923	aminopeptidase, putative, expressed
	irgs	gen	343827	343843	OS03G0819				LOC_Os03g60	Sb01g0033		269985		
3	p	e	59	59	300				470	70	1	3	2701360	glycine-rich protein A3, putative, expressed
	irgs	gen	343862	343881	OS03G0819	Traes_5BL_019DF83A	2541521	2541543	LOC_Os03g60	Sb01g0033		267961		
3	p	e	15	74	400	E.1	04	02	480	40	1	1	2681513	heavy metal-associated domain containing protein, expressed
	irgs	gen	343865	343882	OS03G0819									
3	p	e	71	33	450				None					
	irgs	gen	343945	343956	OS03G0819	Traes_5BL_E86097AA	2507538	2507552	LOC_Os03g60	Sb01g0033		266900		
3	p	e	08	38	600	2.1	35	54	509	30	1	8	2670703	expressed protein
	irgs	gen	343969	344001	OS03G0819	Traes_5BL_F64F43364			LOC_Os03g60	Sb01g0033		266372		
3	p	e	69	84	700	.2			520	20	1	2	2667043	expressed protein
	irgs	gen	344070	344103	OS03G0819	Traes_5BL_47218A29			LOC_Os03g60	Sb01g0033		265497		
3	p	e	93	74	900	C.2			530	10	1	3	2658361	ras-related protein, putative, expressed
	irgs	gen	344128	344191	OS03G0820	Traes_5BL_D526A626	1974720	1974796	LOC_Os03g60	Sb01g0032		264339		
3	p	e	40	39	100	E.2	82	04	550	90	1	9	2652117	ThiF family domain containing protein, putative, expressed
	irgs	gen	344241	344250	OS03G0820	Traes_5BL_D53A846B	2504805	2504815	LOC_Os03g60	Sb01g0032		263690		
3	p	e	80	30	300	E.1	24	44	560	80	1	5	2637456	ZOS3-21 - C2H2 zinc finger protein, expressed
	irgs	gen	344277	344283	OS03G0820	Traes_5BL_8DB8274D	2504590	2504599	LOC_Os03g60					
3	p	e	04	91	400	B1.1	84	54	570					ZOS3-22 - C2H2 zinc finger protein, expressed
	irgs	gen	344304	344313	OS03G0820	Traes_5BL_B35A2E6B	2494432	2494447	LOC_Os03g60	Sb01g0032		262961		
3	p	e	57	64	500	2.1	59	25	580	60	1	8	2630509	actin-depolymerizing factor, putative, expressed
	irgs	gen	344317	344342	OS03G0820	Traes_5BL_F421A6446				Sb01g0032		262662		
3	p	e	00	24	600	.2			None	50	1	0	2628778	
	irgs	gen	344350	344384	OS03G0820	Traes_5BL_5C19F4988			LOC_Os03g60	Sb01g0032		262270		
3	p	e	14	52	700	.3			600	40	1	1	2625591	zinc knuckle domain containing protein, expressed
	irgs	gen	344393	344443	OS03G0820	Traes_5BL_AFDCCFB			LOC_Os03g60	Sb01g0032		261492		
3	p	e	54	87	900	98.1			610	30	1	0	2621046	SEC14 cytosolic factor family protein, putative, expressed
	irgs	gen	344399	344433	OS03G0821									
3	p	e	26	91	000				None					
	irgs	gen	344463	344506	OS03G0821	Traes_5BL_88E594FC			LOC_Os03g60					
3	p	e	27	65	100	F.1			620					DnaK family protein, putative, expressed
	irgs	gen	344487	344501	OS03G0821									
3	p	e	36	83	150				None					
	irgs	gen	344544	344567	OS03G0821	Traes_5BL_7A5C5CC	2555300	2555311	LOC_Os03g60	Sb01g0032		260573		
3	p	e	95	41	200	CB1.1	46	28	630	10	1	3	2607377	dof zinc finger domain containing protein, putative, expressed
	irgs	gen	344642	344650	OS03G0821				LOC_Os03g60					
3	p	e	52	38	250				639					expressed protein
	irgs	gen	344746	344783	OS03G0821	Traes_5BL_F296A638			LOC_Os03g60	Sb01g0031		258154		
3	p	e	97	70	300	F.1			650	90	1	8	2584791	protein phosphatase 2C, putative, expressed
	irgs	gen	344872	344881	OS03G0821									
3	p	e	74	13	350				None					
	irgs	gen	344932	344936	OS03G0821	Traes_5BL_C47B6A9E				Sb01g0031		253916		
3	p	e	35	21	550	D.1			None	60	1	0	2540212	
	irgs	gen	344937	344968	OS03G0821	Traes_5BL_76075E68			LOC_Os03g60	Sb01g0031		253407		
3	p	e	49	57	700	A.1			690	50	1	3	2535896	pentatricopeptide, putative, expressed
	irgs	gen	344982	345012	OS03G0821	Traes_5BL_9F06DE86			LOC_Os03g60	Sb10g0210		462776		
3	p	e	78	91	800	D.1			700	46	10	33	46278849	dolichyl-phosphate beta-glycosyltransferase, putative, expressed
	irgs	gen	345029	345077	OS03G0821	Traes_5BL_C52B6F6D			LOC_Os03g60	Sb01g0031		252063		
3	p	e	81	40	900	6.2			710	30	1	1	2524618	protein kinase domain containing protein, expressed
	irgs	gen	345069	345075	OS03G0821									
3	p	e	62	38	950				None					
	irgs	gen	345075	345112	OS03G0822	Traes_5BL_1133E46E7	2585107	2585136	LOC_Os03g60	Sb01g0031		251509		
3	p	e	97	79	000	.1	15	77	720	20	1	5	2519371	expansin precursor, putative, expressed
	irgs	gen	345089	345108	OS03G0822									
3	p	e	17	58	033				None					
	irgs	gen	345217	345259	OS03G0822	Traes_2AL_F5BCAC9			LOC_Os03g60	Sb05g0062		965350		
3	p	e	48	73	100	4C.1			730	63	5	1	9655063	transposon protein, putative, unclassified, expressed
	irgs	gen	345265	345290	OS03G0822	Traes_5BL_69A0036E			LOC_Os03g60	Sb01g0031		248911		
3	p	e	02	96	200	A.1			740	00	1	9	2491362	expressed protein
	irgs	gen	345292	345314	OS03G0822	Traes_5BL_94BCE486	2568344	2568365	LOC_Os03g60	Sb01g0030		248119		
3	p	e	05	01	300	4.2	17	40	750	90	1	6	2488825	ribosomal RNA large subunit methyltransferase J, putative, expressed
	irgs	gen	345331	345338	OS03G0822	Traes_5BL_1DF722774			LOC_Os03g60	Sb01g0030		247835		
3	p	e	82	28	400	.1			760	80	1	8	2478639	expressed protein
	irgs	gen	345362	345392	OS03G0822	Traes_5BL_C318204D			LOC_Os03g60	Sb01g0030		247367		
3	p	e	80	96	700	2.1			780	70	1	2	2476504	armadillo/beta-catenin-like repeat containing protein, expressed

	irgs	gen	345395	345413	OS03G0822	Traes_5BL_93A15890F	2576775	2576788	LOC_Os03g60	Sb01g0030		246973			
3	p	e	17	16	800	.1	90	82	790	60	1	1	2472656	heat shock protein DnaJ, putative, expressed	
	irgs	gen	345415	345453	OS03G0822	Traes_5BL_1FA8B507	2576797	2576829	LOC_Os03g60	Sb01g0030		246082			
3	p	e	57	09	900	5.1	81	22	800	40	1	2	2463740	transposon protein, putative, unclassified, expressed	
	irgs	gen	345454	345478	OS03G0823	Traes_6BL_7DB20462			LOC_Os03g60	Sb01g0030		244977			
3	p	e	92	56	000	1.1			810	20	1	2	2451851	lectin-like receptor kinase, putative, expressed	
	irgs	gen	345546	345567	OS03G0823	Traes_5BL_72EFF0AF			LOC_Os03g60	Sb01g0030		244193			
3	p	e	49	94	100	0.2			820	10	1	7	2444540	transporter, major facilitator superfamily domain containing protein, expressed	
	irgs	gen	345569	345598	OS03G0823				LOC_Os03g60	Sb01g0030		243960			
3	p	e	68	78	200				820	05	1	4	2440694	transporter, major facilitator superfamily domain containing protein, expressed	
	irgs	gen	345605	345608	OS03G0823				LOC_Os03g60						
3	p	e	98	51	301				820					transporter, major facilitator superfamily domain containing protein, expressed	
	irgs	gen	345623	345624	OS03G0823										
3	p	e	10	79	350				None						
	irgs	gen	345671	345680	OS03G0823	Traes_5BL_8ABF0777			LOC_Os03g60	Sb01g0030		243381			
3	p	e	09	87	400	7.1			840	00	1	0	2436043	BBTI13 - Bowman-Birk type bran trypsin inhibitor precursor, expressed	
	irgs	gen	345743	345769	OS03G0823	Traes_5BL_BB28CAC	2583753	2583778	LOC_Os03g60	Sb01g0029		242655			
3	p	e	35	73	500	E8.1	48	11	850	90	1	3	2429064	peptide transporter PTR2, putative, expressed	
	irgs	gen	345746	345765	OS03G0823										
3	p	e	05	34	550				None						
	irgs	gen	345931	345964	OS03G0823	Traes_5BL_E220C7864	2589639	2589657	LOC_Os03g60	Sb01g0029		241402			
3	p	e	34	87	700	.2	16	55	870	80	1	1	2416789	ras-related protein, putative, expressed	
	irgs	gen	345974	346012	OS03G0823	Traes_5BL_95EF673E3	2599861	2599900	LOC_Os03g60	Sb01g0029		240052			
3	p	e	55	07	800	.1	22	93	880	60	1	6	2403746	nucleobase-ascorbate transporter, putative, expressed	
	irgs	gen	346026	346057	OS03G0823	Traes_5BL_1FC1D39A	2582040	2582079	LOC_Os03g60						
3	p	e	74	56	900	F.2	74	89	890					calmodulin binding protein, putative, expressed	
	irgs	gen	346075	346081	OS03G0824	Traes_1BL_A7523182			LOC_Os03g60	Sb01g0029		238849			
3	p	e	94	74	000	D.1			900	40	1	7	2389294	expressed protein	
	irgs	gen	346100	346138	OS03G0824	Traes_5BL_535609A20	2571276	2571320	LOC_Os03g60	Sb01g0029		238419			
3	p	e	54	89	100	.2	83	05	910	30	1	0	2388209	PPR repeat domain containing protein, putative, expressed	
	irgs	gen	346146	346168	OS03G0824	Traes_5BL_E8E506F82			LOC_Os03g60	Sb01g0029		238123			
3	p	e	10	53	200	.1			920	20	1	0	2383287	Methyltransferase small domain containing protein, expressed	
	irgs	gen	346224	346252	OS03G0824	Traes_5BL_18F56C243	2396404	2396492	LOC_Os03g60	Sb01g0048		386596			
3	p	e	63	04	300	.2	76	37	930	00	1	4	3873753	RNA recognition motif containing protein, putative, expressed	
	irgs	gen	346260	346303	OS03G0824										
3	p	e	25	15	350				None						
	irgs	gen	346307	346332	OS03G0824	Traes_5BL_C9F29F52	1974669	1974707	LOC_Os03g60	Sb01g0029		237684			
3	p	e	76	81	400	F.1	52	61	939	10	1	8	2379525	dolichyl-phosphate beta-glycosyltransferase, putative, expressed	
	irgs	gen	346336	346352	OS03G0824	Traes_5BL_0D76D563	2568720	2568749	LOC_Os03g60	Sb01g0028		236085			
3	p	e	40	44	500	F.1	24	65	950	80	1	2	2362686	expressed protein	
	irgs	gen	346368	346384	OS03G0824	Traes_5BL_70975BD6	2602105	2602127	LOC_Os03g60	Sb01g0028		237134			
3	p	e	93	52	600	2.2	02	09	960	90	1	5	2372839	cytokinin-N-glucosyltransferase 1, putative, expressed	
	irgs	gen	346371	346383	OS03G0824										
3	p	e	90	51	650				None						
	irgs	gen	346633	346682	OS03G0825	Traes_5BL_06912B45	2579124	2579193	LOC_Os03g61	Sb09g0188		472201			
3	p	e	75	23	300	B.1	89	61	010	35	9	13	47221123	protein kinase family protein, putative, expressed	
	irgs	gen	346690	346711	OS03G0825	Traes_5BL_68C95B39	2579205	2579229	LOC_Os03g61	Sb01g0028		232574		mitochondrial import inner membrane translocase subunit Tim, putative,	
3	p	e	05	41	400	6.1	72	11	019	40	1	1	2327731	expressed	
	irgs	gen	346713	346744	OS03G0825	Traes_5BL_F7FF5F98	2595524	2595556	LOC_Os03g61	Sb01g0028		232307			
3	p	e	02	91	500	C.2	18	39	030	30	1	8	2325424	transcription termination factor nusG family protein, expressed	
	irgs	gen	346749	346769	OS03G0825	Traes_5BL_2EFE0C7E			LOC_Os03g61	Sb01g0028		231977			
3	p	e	79	23	600	31.1			040	20	1	4	2321281	GIL1, putative, expressed	
	irgs	gen	346800	346879	OS03G0825	Traes_5BL_CC831762	2562418	2562491	LOC_Os03g61	Sb01g0028		230820			
3	p	e	09	04	700	7.1	34	45	050	10	1	9	2315871	FG-GAP repeat-containing protein, putative, expressed	
	irgs	gen	346882	346906	OS03G0825	Traes_5BL_DA4CA26	2563006	2563033	LOC_Os03g61	Sb01g0028		230571			
3	p	e	87	76	800	BA.2	33	07	060	00	1	2	2308143	protein kinase domain containing protein, expressed	
	irgs	gen	346885	346909	OS03G0825										
3	p	e	62	11	850				None						
	irgs	gen	346929	346961	OS03G0825				LOC_Os03g61						
3	p	e	11	98	900				070					expressed protein	
	irgs	gen	347008	347054	OS03G0826	Traes_4AL_09CB42FF			LOC_Os03g61	Sb08g0076		138668			
3	p	e	40	14	000	B.2			080	40	8	18	13872324	expressed protein	
	irgs	gen	347009	347044	OS03G0826										
3	p	e	76	60	100				None						
	irgs	gen	347070	347077	OS03G0826				LOC_Os03g61						
3	p	e	24	36	200				090					expressed protein	
	irgs	gen	347081	347113	OS03G0826	Traes_5BL_F3404AA5	2592561	2592590	LOC_Os03g61	Sb01g0027		228625			
3	p	e	03	23	300	6.1	62	69	100	75	1	4	2289248	GDP-mannose transporter, putative, expressed	

	irgs	gen	347127	347150	OS03G0826	Traes_5BL_72FA8AB3	2597462	2597496	LOC_Os03g61	Sb01g0027		228027		
3	p	e	97	06	400	1.1	48	07	110	70	1	9	2282517	RNA-binding motif protein, putative, expressed
	irgs	gen	347155	347210	OS03G0826	Traes_5BL_171EB1A0	2597497	2597536	LOC_Os03g61	Sb01g0027		227202		
3	p	e	21	72	500	1.1	14	90	120	60	1	3	2277198	anthranilate synthase component I-1, chloroplast precursor, putative, expressed
	irgs	gen	347243	347272	OS03G0826	Traes_5BL_F6BC648F	2595573	2595601	LOC_Os03g61	Sb01g0027		226270		
3	p	e	28	41	600	3.1	32	53	130	50	1	7	2265582	phosphoesterase family protein, putative, expressed
	irgs	gen	347282	347315	OS03G0826	Traes_5BL_2F346D36	2594757	2594785	LOC_Os03g61	Sb01g0027		224098		
3	p	e	21	29	700	E.1	60	62	140	10	1	0	2243493	expressed protein
	irgs	gen	347334	347347	OS03G0826	Traes_4AL_D4B405C			LOC_Os03g61	Sb01g0027		224425		
3	p	e	90	06	800	DE.1			150	20	1	3	2245387	expressed protein
	irgs	gen	347390	347405	OS03G0826				LOC_Os03g61					
3	p	e	11	04	900				160					expressed protein
	irgs	gen	347560	347587	OS03G0827	Traes_5BL_45662533A	2576960	2576985	LOC_Os03g61	Sb01g0026		223511		
3	p	e	11	67	500	.1	46	27	200	90	1	2	2237901	expressed protein
	irgs	gen	347592	347629	OS03G0827	Traes_5BL_11818F347.	2601790	2601825	LOC_Os03g61	Sb01g0026		223032		
3	p	e	90	82	600	1	51	88	210	80	1	1	2234081	domain of unknown function domain containing protein, expressed
	irgs	gen	347659	347722	OS03G0827	Traes_5BL_BD895CF8	2601938	2601967	LOC_Os03g61	Sb01g0026		216996		
3	p	e	40	84	700	4.2	19	85	220	60	1	2	2176738	DEAD-box ATP-dependent RNA helicase 3, putative, expressed
	irgs	gen	347786	347794	OS03G0827				LOC_Os03g61	Sb01g0026		216051		
3	p	e	69	67	900				240	50	1	6	2161536	expressed protein
	irgs	gen	347854	347872	OS03G0828	Traes_5BL_FA3E0DB0	2588617	2588641	LOC_Os03g61	Sb01g0026		212841		
3	p	e	73	45	100	3.1	29	17	260	10	1	5	2130430	ribosomal L18p/L5e family protein, putative, expressed
	irgs	gen	347920	347938	OS03G0828	Traes_5BL_1512240F3.	2602311	2602333	LOC_Os03g61	Sb01g0026		213655		
3	p	e	25	42	300	1	93	14	270	30	1	5	2139282	OsMan04 - Endo-Beta-Mannanase, expressed
	irgs	gen	347983	348007	OS03G0828	Traes_5BL_1512240F3.	2602311	2602333	LOC_Os03g61	Sb01g0026		213655		
3	p	e	12	60	500	1	93	14	280	30	1	5	2139282	OsMan05 - Endo-Beta-Mannanase, expressed
	irgs	gen	348057	348076	OS03G0828	Traes_5DS_D50F61EE			LOC_Os03g61	Sb01g0026		214431		
3	p	e	11	00	600	5.1			290	40	1	5	2149695	ATCHX, putative, expressed
	irgs	gen	348100	348132	OS03G0828	Traes_4AL_C74A4928			LOC_Os03g61					
3	p	e	77	40	800	0.1			310					receptor-like protein kinase, putative, expressed
	irgs	gen	348171	348203	OS03G0829	Traes_5BL_3AF852D9	2604782	2604827	LOC_Os03g61	Sb01g0115		103985		
3	p	e	25	70	000	A.1	04	56	330	70	1	25	10401645	fumarylacetoacetase, putative, expressed
	irgs	gen	348206	348235	OS03G0829	Traes_5BL_52DC7F34	2623350	2623354	LOC_Os03g61	Sb01g0024		196586		
3	p	e	87	29	100	4.1	81	98	340	10	1	2	1968700	hydrolase, alpha/beta fold family domain containing protein, expressed
	irgs	gen	348255	348289	OS03G0829	Traes_3B_50DC6683B.			LOC_Os03g61	Sb01g0024		196023		
3	p	e	02	44	200	1			360	00	1	2	1965330	hydrolase, alpha/beta fold family domain containing protein, expressed
	irgs	gen	348260	348262	OS03G0829									
3	p	e	12	69	333				None					
	irgs	gen	348262	348302	OS03G0829									
3	p	e	76	88	466				None					

List of abbreviations

ANOVA	Analysis of variance
bp	base pairs
CHA	Chemical hybridization agents
CIMMYT	International Maize and Wheat Improvement Center
cM	centi Morgan
CMS	Cytoplasmic male sterility
DArT	Diversity array technology
DNA	Deoxyribonucleic acid
ExHet	Expected heterozygosity
FAO	Food and agriculture organization of the United Nations
GBS	Genotyping by sequencing
GEBV	Genomic estimated breeding value
GS	Genomic selection
GWAS	Genome-wide association study
Ha	Tectare
HD	Heading date
IBS	Identity-by-state
InDel	Insertion and delition polymorphisms
LD	Linkage disequilibrium
LDM	Leaf dry mass
LOESS	Locally estimated scatterplot smoothing
M	Million
MAF	Minor allele frequency
MAS	Marker-assisted selection
Mbp	Mega base pairs
MRD	Modified rogers distance
NAM	Nested association mapping
PCA	Principal component analysis
QTL	Quantitative trait locus
R/S	Root-to-shoot ratio
RDM	Root dry mass
RL	Root length
SL	Shoot length
SNP	Single-nucleotide-polymorphism
t	Tons
UPGMA	Unweighted pair group method with arithmetic mean

Declaration

I declare that the dissertation here submitted is entirely my own work, written without any illegitimate help by any third party and solely with materials as indicated in the dissertation. I have indicated in the text where I have used texts from already published sources, either word for word or in substance, and where I have made statements based on oral information given to me. At all times during the investigations carried out by me and described in the dissertation, I have followed the principles of good scientific practice as defined in the "Statutes of the Justus Liebig University Gießen for the Safeguarding of Good Scientific Practice".

Acknowledgments

There are many different people who have supported me during my work for this dissertation and who deserve my grateful thankfulness for everything they did.

First of all, I would like to express my deep gratitude to my first PhD supervisor and head of the Department of Plant Breeding at the JLU Giessen, Prof. Dr. Rod J. Snowdon. I could not have been luckier to find a boss like him. The way Rod has supervised me throughout the last years was unprecedented and characterized by endless patience, trust and scientific professionalism. Being an exotic person, as a wheat scientist in a leading rapeseed research institute, he still gave me all the support in the world when I needed it. Besides his impressive scientific competence and his untiring dedication for plant breeding research I was especially impressed and inspired by his engaging personality, his exemplary open-mindedness, creativeness and the way how he manages stressful and difficult situations. Moreover, he made it possible for me to visit many of the world's most exciting research events and establishments which tremendously broadened both, my scientific and personal horizon - I've never taken this for granted! Rod, thank you so much for the countless fruitful scientific discussions we've had, for the encouraging way how you supervised me, for your patient, professional scientific assistance and also of course for the nice beer tastings (you and your son are excellent beer brewers!). I hope to be able to work together with you for many more years!

I also wish to express my sincere thanks to my second supervisor and head of the department for Biometry and Population Genetics, Prof. Dr. Matthias Frisch. You were always there for me whenever I needed assistance with statistics or genetics and especially gave me so much invaluable input for my first publication, at a time when I was an absolute beginner and didn't know where to start with. The way how you deal with scientific problems, how you convert complex scenarios that often took me weeks to get my head around into clear and transparent facts was extremely instructive and exemplary. Your practical, organized and solution-oriented way to solve problems as well as your great effectiveness is very inspiring, and the couple of legendary IFZ parties we've had will remain unforgotten!

Another very important reference person for my scientific career I would like to gratefully thank is Prof. em. Dr. Wolfgang Friedt. Since the first time I met him when he was still the head of the Department of Plant Breeding at the JLU in Giessen, I've always been very impressed by his

whole appearance and personality, his extensive scientific and practical knowledge and the incomparable achievements he made for plant breeding through his great career. He was the one who evoked my interest in plant breeding research during my very first Bachelor semester and subsequently gave great support, trust, assistance and care during my whole studies, including my Bachelor and Master theses. Thank you very much for always being such a great inspiration for so many years until today now, for sharing your great knowledge with me, and also thank you very much for the invaluable input you are still giving in regular discussions in our department.

I also want to thank all my office mates and colleagues that made my everyday-life so pleasant and productive. Especially the daily scientific discussions and trips to conferences with Christian Werner were really productive and fruitful. Thank you very much for improving my stats knowledge so much in our daily (sometimes heated) discussions, for being such a good friend and for always sharing your hyper-critical view on anything. Further, I'm very thankful for the nice past years I've had with my "Chinese Prof." and good friend Lunwen Qian. Discussing new research with him was always extremely helpful and he gave me great input during all my work. Thanks also for serving as my personal impact factor app, whenever I was unsure about the exact ranking of a certain journal. Further, I would like to thank Stefanie Lück for patiently supporting me during the beginning of my studies and for becoming such a good friend. Many more thanks I also want to send out to Habib Jan for discussions and the fun times we had in the office together with our "Profess", and even more go out to Andreas Stahl, Benjamin Wittkopp, Christian Obermeier, Sarah Hatzig, Sarah Schießl, Birgit Samans, Timm Bernhard, Steffen Windpassinger, Iulian Gabur, Fabian Grandke, Peter Krause, Roman Gäbelein and Harmeet Singh Chawla for being nice colleagues, as well as Sven Gottwald who was mainly involved in the proposal of my PhD project. Many thanks also to my co-authors Gabriel Keeble-Gagnère and Rudi Appels from Australia, Sebastian Parra-Londono and Ralf Uptmoor from the University of Rostock and my project partners Stefan Kontowski (W. von Borries-Eckendorf GmbH & Co. KG.) and Bertrand Schuiling (Wiersum Plantbreeding BV) for the nice cooperation.

I would further like to express my thankfulness to all the people that realized the "real work" of my dissertation. First of all, many, many thanks to my most-favorite Greek lab-officer Stavros Tzigos. Thanks for being such a professional and well-organized colleague and at least as important: thank you for being such a good friend. My everyday-life in the institute wouldn't have been half as enjoyable without you. Also many thanks to Andreas Welke who provided

tireless assistance during the most extreme work in the lab and greenhouse. Your quiet and unwavering demeanor was the exact right counterpart for me and without you half of my thesis would not have been possible - I'm truly thankful for that. Further I would like to thank my other great colleagues Annette Plank, Birgit Keiner, Nelly Weis, Liane Renno, Sebastian Brinker, Lisa Unterberg, Swetlana Renner, Petra Kretschmer and Burkhard Lather for their unequalled and professional commitment during the many greenhouse and lab experiments! I could always 110%ly count on you, thank you so much! I would also like to thank my friends Alexander Grüner and Helge Flüß who always had an open ear and discussed my research throughout my whole studies.

Further, I also want to thank the centerpiece of our department, the secretary. Sabine Schomber and Ulla Riedmeier took care of all the organizational and formal things in an exemplary and professional way and ensured that I can exclusively focus on my scientific work. I really appreciate that and was very happy to have you!

Many more thanks go out to my family and friends that backed me up in any conceivable way. Above all Freya who supported me with her unconditional love and understanding and gave fresh heart to me in all the frustrating situations that came up along my way. She provided the best and most recreative private environment that one could ever wish for; she was partner, manager, psychotherapist and best friend at the same time and kept my back free from anything so that I could exclusively focus on my work. I'm truly and deeply thankful for everything you did for me and have never taken this for granted!

My final sincere thanks go out to my beloved family. Firstly, to my mother Barbara and father Peter Voss-Fels who initially encouraged me to start with the PhD and always backed me up. You both have always been exemplary for me and making you proud was one of my major drivers during my work. I'm also really thankful to have such amazing brothers, Torben, Finn and Kim – you mean everything to me!

The work was funded by the German Federal Ministry of Food and Agriculture (BMEL), grant number FNR-22408212.