**High-throughput SNP genotyping in modern and wild emmer crosses for drought tolerance using a combined association and linkage analysis**

Stuart J. Lucas1\*, Ayten Salantur2, Selami Yazar2, Hikmet Budak3,,4\*

1SU Nanotechnology Research and Application Centre, Sabanci University, Turkey

2Breeding and Genetics, Field Crops Central Research Institute, Turkey

3Faculty of Engineering and Natural Sciences, Sabanci University, Turkey

4Department of Plant Sciences and Plant Pathology, University of Montana, USA

\*Corresponding author. E-mail: [hikmet.budak@montana.edu](mailto:hikmet.budak@montana.edu)

412 Leon Johnson Hall,

Department of Plant Sciences and Plant Pathology,

Montana State University, Bozeman, MT, USA

Email: [slucas@sabanciuniv.edu](mailto:slucas@sabanciuniv.edu)

Sabanci University, SUNUM 1110

Üniversite Cad. 27, Orhanlı

34956 Tuzla, İstanbul

**Abstract:** Tetraploid durum wheat (*Triticum durum* Desf.) is a major world crop that is grown primarily in areas of the world that experience periodic drought, and therefore breeding durum wheat for drought tolerance is a priority. High-throughput single-nucleotide polymorphism (SNP) genotyping techniques have greatly increased the power of linkage and association mapping analyses for bread wheat, but as yet there is no durum-wheat specific platform available. In this study, we evaluate a the new 384HT Wheat Breeders Array for its usefulness in tetraploid wheat breeding by genotyping a breeding population of F6 hybrids, derived from multiple crosses between *T. durum* cultivars and wild and cultivated emmer wheat accessions. Using a combined linkage and association mapping approach, we generated a genetic map including 1345 SNP markers, and identified markers linked to 6 QTLs for coleoptile length (2), heading date (1), anthocyanin accumulation (1) and osmotic stress tolerance (2). We also developed a straightforward approach for combining genetic data from multiple families of limited size that will be useful for evaluating and mapping pre-existing breeding material.

**Keyword index:** *Triticum durum;* SNP genotyping; linkage analysis; association mapping; osmotic stress tolerance; wheat wild relatives.

**Introduction**

Durum wheat (*Triticum durum* Desf.) is the second most cultivated wheat species worldwide after bread wheat (*Triticum aestivum* L.), with current global production of ~38 million tons annually (International Grains Council 2016). *T. durum* is extensively cultivated in the Meditarranean basin under rain-fed conditions, where water availability is a major limiting factor for yield (Giunta, Motzo, and Deidda 1993; Araus et al. 2003). Therefore, with future climate change expected to increase the incidence of extreme climatic events in this region (Challinor et al. 2014), breeding durum wheat varieties with greater abiotic stress tolerance is a high priority.

Cultivated *T. durum* has a tetraploid (AABB) genome and is thought to have been developed by a lengthy selection process from wild emmer wheat, *Triticum dicoccoides* (Körn. ex Asch. & Graebner) Schweinf., with cultivated emmer wheat (*Triticum dicoccon* Schrank) representing an early step in this process. Wild relatives of cultivated wheat harbour valuable genetic resources that were eliminated during the domestication process (Tanksley and McCouch 1997), including stress tolerance traits. For example, a screen of 200 *T. dicoccoides* accessions identified genotypes with significantly better tolerance of water deficiency than modern *T. durum* cultivars (Ergen and Budak 2009). Therefore, introgression of stress tolerance loci from emmer wheat is a promising approach for breeding climate resilient durum wheat.

Tolerance of abiotic stresses such as drought is a complex characteristic involving multiple genes and traits that make different contributions depending on the precise environmental conditions (Tardieu 2012). As a result, efforts to breed for abiotic stress tolerance using phenotypic data alone or using low-density molecular markers have met with limited success. High-throughput genotyping using Single Nucleotide Polymorphisms (SNPs) has the potential to overcome this limitation, as SNPs cover the genome at much higher density than other marker types (Bérard et al. 2009); in theory there should be one or more SNPs in every functional allele, enabling highly complex traits to be dissected. In recent years, considerable effort has been invested in mapping and sequencing the genome of bread wheat (Brenchley et al. 2012; Lucas et al. 2014; International Wheat Genome Sequencing Consortium 2014), from which 100,000s of *T. aestivum* SNPs have been identified (Allen et al. 2012; Lorenc et al. 2012). Furthermore, microarrays utilizing these SNPs for high-throughput genotyping have been developed for both the Illumina iSelect (Wang et al. 2014b) and Affymetrix Axiom (Winfield et al. 2016) platforms. In durum wheat, several thousand SNPs have been identified *de novo* (Trebbi et al. 2011; van Poecke et al. 2013) but there is no *T. durum* specific genotyping platform available. However, a subset of the SNPs present on each of the aforementioned *T. aestivum* genotyping arrays have also been shown to be informative for genetic mapping in some *T. durum* populations (Maccaferri et al. 2015; Winfield et al. 2016).

Genetic studies aiming to dissect quantitative traits such as stress tolerance use two strategies to identify Quantitative Trait Loci (QTLs). Linkage analysis in biparental mapping populations is traditionally used to map QTLs, but suffers from limited resolution, particularly in regions of the genome with low recombination frequencies; also, the mapping population only contains as much genetic diversity as the parental lines. In contrast, in association mapping methods a panel of genetically diverse lines is genotyped using a large number of markers to identify marker alleles that are statistically associated with a trait. This can give much higher genetic resolution than linkage analysis, but unknown population structure within the mapping panel can lead to false positive associations, while genuine trait alleles may be overlooked if they are rare in the diversity panel (Ott, Wang, and Leal 2015). Therefore, several population designs have been proposed to combine the complementary strengths of linkage analysis and association mapping, such as Multi-parental Advanced Generational InterCross (MAGIC) (Holland 2015) and Nested Association Mapping (NAM) (McMullen et al. 2009). While undoubtedly powerful, these populations are expensive and labour-intensive to produce. Therefore, there is a need for more general methods that can combine the benefits of linkage analysis and association mapping in populations from existing plant breeding programmes.

In this study, we aimed to characterize a diverse population of hybrids generated from crosses between Turkish *T. durum* cultivars and wild tetraploid accessions. We assessed the utility of the Axiom 384HT Wheat Breeders’ Genotyping Array for tetraploid wheats, as it offers higher throughput and lower cost per sample than any other currently available genotyping platform, but is primarily designed for use with elite bread wheat lines. In addition to field phenotypic observations, we studied the development of young seedlings under osmotic stress, as osmotic stress forms a component of multiple abiotic stress conditions such as drought and cold. From the genotyping data, we aimed to identify SNP markers associated with stress tolerance traits using a combination of linkage and association mapping.

**Materials and Methods**

**Plant material and breeding**

Seeds for all *T. durum* cultivars, breeding lines and wild relatives were from those maintained in the gene bank of the Field Crops Central Research Institute, Ankara, Turkey. As part of an ongoing breeding program, 12 different Turkish *T. durum* cultivars were used as pollen donors to fertilize 19 tetraploid relatives originally collected in Turkey (14 accessions of *T. dicoccoides*, 4 *T. dicoccon*, and 1 *T. araraticum*). Successful hybridizations were back-crossed once to the *T. durum* parent, and then propagated by self-fertilization for 5generations. From the F5 hybrids, 500 individuals were selected for preliminary yield trials at the F6 generation, and were planted in plots at a single field location in Ankara province. During the trials, phenotypic measurements including coleoptile length, days to heading, days to maturity and plant height were recorded for 10 individual plants in each plot, while assessments of biomass, lodging, anthocyanin (visible through reddening of stalks and/or heads), morphology of heads and seeds, and yield were recorded for each plot as a whole. 100 representative lines incorporating the phenotypic diversity observed in the population were then selected for genotyping. The cultivars and accessions used and the parentage of the F6 hybrids used in this study along with phenotypic data are summarized in Table S1.

**DNA isolation**

Single spikes were obtained for each of the selected F6 hybrids and DNA was isolated from individual seeds. After weighing, seeds were placed in 2ml tubes with a 5mm steel bead and frozen at -80oC for 20 minutes before bead beating in a Tissue Lyser II (Qiagen, Düsseldorf, Germany) at 30 s-1 for 1 minute. If necessary the freezing and beating was repeated until the sample was fully homogenized. DNA was then isolated using the Nucleospin Plant II kit (Macherey-Nagel, Düren, Germany). The following modifications to the manufacturer’s protocol were empirically determined to improve yield and purity: lysis buffer PL2 (SDS-based) was used rather than PL1; lysis incubation at 65oC was extended to 30 minutes; after initial lysis, the samples were centrifuged at 11,000 rcf for 5 minutes, and the cleared lysate loaded on to the filtration column; final elution was carried out using 2 x 30 µl Elution Buffer (5 mM Tris-HCl, pH 8.5) and both eluates combined. DNA yield and purity was estimated using a Nanodrop 2000c Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) and the integrity of the isolated DNA confirmed by agarose gel electrophoresis.

**SNP Genotyping**

Accurate concentrations of high molecular weight DNA were measured using the Quant-iT PicoGreen dsDNA Assay kit (ThermoFisher Scientific, Waltham, MA, USA) according to the manufacturer’s instructions, and 1.5 µg of gDNA from each of the parental accessions and 100 F6 hybrids were diluted into a final volume of 50 µl using 10 mM Tris-HCl ph 8.0. The samples were processed and hybridized in duplicate to the Axiom 384HT Wheat Breeder’s Genotyping Array (Affymetrix, Santa Clara, CA, USA) at the Bristol Genomics Facility (Bristol University, UK). Quality control and SNP calling was carried out with Axiom Analysis Suite v1.1.0.616 software (Affymetrix) in ‘Best Practices Workflow’ mode. In order to include the wild relatives in the analysis the default quality control parameters were relaxed to the following values: DQC cutoff = 0.795, QC call rate cutoff = 92%, SNP call rate cutoff = 92%. The *T. durum* parentallines and the great majority of F6 hybrids exhibited a QC call rate > 98%. In order to take the low heterozygosity of the inbred lines into account during genotype calling, an inbred\_het\_penalty value of 7 was specified for all the *T. durum* parents and hybrid lines, but not the wild tetraploids. The accuracy of the automatic SNP calls was checked both using follow up scripts (PS Supplemental, OTV Caller) and manually as described in the results section.

**Selection of SNPs for Linkage Analysis**

Genotype calls were exported to a text file for all polymorphic SNPs that passed quality filters. Custom Perl scripts were used to combine genotyping results from duplicate samples and select the most informative SNPs prior to linkage analysis. Firstly, genotypes were re-coded using the available pedigree information as follows: 0 = homozygote matching *T. durum* parent, 2 = homozygote matching wild tetraploid parent, 1 = heterozygote, X = no call or uninformative (both parents of a particular hybrid have the same genotype). All SNPs with informative calls in fewer than 20 lines were eliminated, and the remaining 7714 SNPs were binned to eliminate redundant markers. In total 678 bins of 2 or more SNPs with identical genotyping results across all lines were identified; a single representative of each bin was retained and 1188 redundant SNPs removed from the linkage analysis. Next, SNPs that showed significant segregation distortion were identified and eliminated using Chi-squared tests with sequential Bonferroni correction (Rice 1989). The remainder were sorted by percentage of informative calls (highest to lowest) and all that were informative in 70% or more of the F6 hybrids were used to construct the framework linkage map.

**Linkage Map Construction and Comparison**

Genetic map construction was carried out using the multi-point maximum likelihood approach implemented in CarthaGene v.1.2.3 (de Givry et al. 2005). Linkage groups were evaluated using SNPs that were also located on the durum wheat and bread wheat consensus linkage maps (Maccaferri et al. 2015)(Winfield et al. 2016). Correspondences between SNPs used on the Axiom arrays and the Infinium iSelect array used by Maccaferri were obtained from CerealsDB (<http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/axiom_download.php>, last updated 31.10.16). Firstly, markers with identical genotypes across all individuals were merged into single loci to simplify the mapping calculation. Initial grouping of markers was carried out using stringent limits of minimum LOD (logarithm of the odds) = 8, maximum *r* (recombination frequency) = 0.3, after which each linkage group contained markers from a single chromosome. Linkage groups containing markers from the same chromosome were then selected and combined by re-grouping at LOD = 3 or progressively lower limits. Markers within each group were ordered using Carthagene’s ‘build’ algorithm, followed by optimizing the order with ‘flips’ and ‘polish’, selecting the order with the maximum likelihood after each step. After this round of mapping, all linkage groups consisted of well-ordered blocks of markers, but in some groups the arrangement of these blocks relative to each other could be varied with minimal change in the likelihood calculation. Therefore, the linkage groups were aligned to the aforementioned consensus maps using MapDisto v2.0 (Lorieux 2012), and the matrices of 2-point LOD and *r* values within each group visualized. The order of blocks most consistent with the consensus maps was used as the starting point for a second round of mapping and order optimization. Finally, individual markers that were poorly linked to the rest of the group were eliminated using the ‘squeeze’ command. Genetic map distances were calculated using Kosambi’s mapping function (Kosambi 1943) after each step. SNP markers included in the linkage maps were mapped to contigs from the draft bread wheat genome sequence (International Wheat Genome Sequencing Consortium 2014) using standalone BLAST v2.2.30 (Camacho et al. 2009). Alignments considered as hits included the full query sequence (60-70nt flanking each SNP) with 3 or fewer mismatches and no gaps. Linkage maps were visualized using MapChart v2.3 (Voorrips 2002).

**Root morphology analysis**

The root morphology of seedlings under osmotic stress was assessed on plant agar plates as previously described (Verslues et al. 2006). Briefly, 100mm petri dishes were filled with 20ml of 15 g/l micropropagation agar dissolved in half strength Murashige & Skoog (MS) medium with 6 mM MES buffer. After solidifying, these were overlaid with 30ml of solutions of PEG-6000 (Polyethyleneglycol, average Mr = 6000) dissolved in the same medium and left overnight to allow the PEG to diffuse into the agar. The overlay solution was then discarded and the plates used immediately. During preliminary experiments using seeds of *T. durum* cultivar Kiziltan-91, a small amount of PEG (5% w/v in the overlay) was observed to stimulate germination compared to plates containing no PEG, while an overlay of 40% (w/v) PEG resulted in an average 30-50% reduction in subsequent root growth.

Therefore, from each F6 hybrid, 10 seeds were surface washed with 0.5% sodium hypochlorite solution, and then sown on to a plate infused with 5% PEG-6000 to germinate for 24 hr. Next, 5 seeds each were transferred to plates infused either with 40% PEG-6000 (osmotic stress) or MS/MES medium only (control). These plates were incubated in the dark at room temperature, and morphology measurements taken 4 and 6 days after initial sowing. Seeds that failed to develop were discarded; all seedlings were lifted from the agar, washed, and imaged using an Epson Perfection V700 Photo scanner. Images were analyzed with WinRHIZO software (Regent Instruments, Quebec, Canada). Data for 5 individuals under each condition were processed and analyzed in Microsoft Excel. Statistical assessment of the effects of genotype and treatment was carried out using 2-way ANOVA with the random model, as implemented in the Real Statistics Resource pack for Excel (www.real-statistics.com).

**Association Mapping**

Association analysis was carried out using TASSEL v5.0 software (Bradbury et al. 2007). Genotype base calls for all SNPs were converted into PLINK format for import into TASSEL, using the genetic map generated above to order the SNPs. The phenotype data used included 10 biological replicates (individual plants) for the field morphological traits, and 5 biological replicates (individual seedlings) for the osmotic stress experiments. The genotyping data were filtered to exclude loci with a minor allele frequency < 0.05. Underlying population structure was tested using Principal Component Analysis, and a kinship matrix calculated from the genotyping data using the Centered\_IBS equation (Endelman, Jannink, and Holley 2012). Marker-trait associations detected using the Mixed Linear Model with default options. Significance thresholds for marker-trait associations were selected by examination of QQ plots for each trait (Figure S1), as the value above which observed p-values were higher than expected p-values (p < 0.001 in all cases). Where SNPs adjacent to one(s) that passed this threshold also had p < 0.05, they were included with it as a cluster of SNPs associated with the trait of interest. The MLM\_statistics and MLM\_effects tables produced by TASSEL were exported to text files, and the SNPs and genotypes associated with traits extracted in Microsoft Excel.

**Results**

**Genotyping a complex tetraploid wheat population**

As part of the ongoing durum wheat breeding programme at the Field Crops Central Research Institute, Ankara, 12 Turkish durum wheat cultivars and breeding lines had been hybridized with 15 wild and 4 cultivated emmer wheat accessions in an incomplete factorial design, and propagated to the F6 generation (Materials and Methods). Following preliminary yield trials, 100 F6 hybrids with diverse parentage and phenotypes were selected for use in this study (summarized in Table S1).

DNA was isolated from individual seeds and all the F6 hybrids were genotyped in duplicate on the Wheat Breeders Genotyping array, along with all the parental lines. As this array was designed primarily for hexaploid wheat cultivars and landraces, quality control cut-offs were relaxed to accommodate the genetic diversity and absence of the D sub-genome in our population, as described in Materials and Methods. In fact, for all *T. durum* and the great majority of F6 hybrid samples, positive calls were returned for > 97.5% of the array SNPs (Fig. 1A). As expected for inbred lines, ≤ 2% of SNP calls in these samples were heterozygous. In contrast, the emmer wheat lines had slightly lower SNP call rates and a higher proportion of heterozygous markers (3-8%). A further 6 F6 hybrid and 3 *T. dicoccoides* samples had markedly lower call rates and more than twice the number of heterozygous calls than other samples of the same type; these probably represent mixed or poor quality samples, and so were manually excluded from genotype calling. This eliminated one F6 hybrid (AS454), but for all the others a higher quality duplicate sample was available. Genotype calls were compared for all pairs of good quality replicates, and 99.59 ± 0.28% of calls were identical within each pair, demonstrating the reproducibility of the system.

The high call rate was unexpected, given that approximately 1/3 of the SNPs on the array were mapped to the D genome, which is absent in our tetraploid samples. Therefore, the classes of genotype calls were compared between the 3 sub-genomes (Table 1). Although the average positive call rate (~98.8%) showed no difference between SNPs from each sub-genome, a much smaller proportion of SNPs from the D genome were found to be polymorphic in our population. This could be explained by similar, but non-polymorphic, loci from the A and B genomes binding to probes for D genome SNPs. However, a substantial number of SNPs mapped to the D genome were also polymorphic in tetraploid wheats, suggesting that these SNPs have homoeologs on the A or B genomes.

The Axiom Analysis software assigns genotypes to each SNP by clustering the signals from all samples based on the sum of (Size) and difference between (Contrast) the 2 coloured fluorescence signals corresponding to the 2 alleles for each SNP (Fig. 1B-1F). In polyploids, additional signals from homoeologous sequences can confuse the clustering algorithm; therefore, all of the cluster plots for polymorphic SNPs were checked visually. Fig. 1B shows a symmetrical arrangement with the heterozygote cluster at zero contrast, which is observed for SNPs that only map to one of the sub-genomes and so are effectively diploid. Fig. 1C shows an asymmetrical arrangement, in which one of the homozygous clusters is closest to zero contrast; this is caused by background signal from a homoeologous non-polymorphic locus (e.g. the SNP shown here was genetically mapped to chr. 4B, but its probe sequence also aligns to a sequence contig from chr. 4AS). Therefore, results such as those in Fig. 1D which contain only 2 clusters are ambiguous. The smaller cluster containing the 4 *T. dicoccon* and several hybrid samples has been called as heterozygous, as in Fig. 1B. However, given the low heterozygosity in our population, it is more likely that this cluster is homozygous and the (empty) heterozygous cluster should be in between the 2 observed clusters, as in Fig. 1C. Ambiguous SNPs such as this were classified as ‘NoMinorHom’ and were not used in the following analyses. Another possible source of incorrect calls is ‘Off Target Variants’ (OTVs), which occur in samples that have a secondary mutation in the SNP flanking region so that they do not hybridize to the array probes. Fig. 1E shows an example where one of the *T. durum* parental lines and several of its descendants have an OTV, creating a 4th cluster with lower signal intensity below the heterozygous cluster. This example was correctly called as an OTV by the ‘OTV caller’ script, but others were not (e.g. Fig 1F, where the OTVs are mistakenly included in the AA cluster), particularly when the clusters were asymmetrical. In these cases, the genotype calls of the OTV cluster were manually corrected.

In total, the clustering algorithm classified 9113 SNPs (25.9% of all on the array) as ‘PolyHighRes’ (polymorphic, high resolution). From these, 508 were discarded for having a minor allele frequency < 0.04, 14 were eliminated due to poor clustering, and 234 had some of their genotype calls manually corrected to OTV. Furthermore, SNPs classified as OTV or ‘high variance’ were also checked manually, and 893 of these were considered to give reliable genotype calls. With these included, 9484 polymorphic SNPs were used for the subsequent analyses.

**Genetic linkage mapping in a complex population**

In crop plants, genetic linkage maps are generally produced from large, biparental populations. In contrast, in humans and other animals where such large families are unavailable, observed rates of recombination in multiple families with are combined to estimate genetic linkage. Similarly, we developed an approach using the multiple crosses represented by the F6 hybrids used in this study to generate a consensus linkage map (described in detail in Materials and Methods). Initially, genotype calls for each hybrid were compared with those of its parental lines, and re-coded using a common scheme of 0 = *T. durum* parental homozygous genotype, 2 = emmer wheat parental homozygous genotype, 1 = heterozygous, X = missing data. This enabled direct comparison between families, and allowed OTVs to be included in the analysis as one of the parental genotypes. In families where both parents had the same genotype for a particular SNP, all F6 progeny were re-coded as ‘missing data’ for that SNP, as any recombinations would not be detected. After eliminating markers that were redundant, showed segregation distortion, or had >30% missing data, 1782 SNPs were used to construct the linkage map. A bread wheat consensus map that had previously been constructed using Axiom SNPs (Winfield et al. 2016) was used to assign the linkage groups to chromosomes; only 16 of the markers used were previously mapped to the D genome.

The final linkage map included 1346 SNP markers representing 931 unique genetic loci, with a total map length of 3972 cM (Figure 2, Table 2). The 14 linkage groups corresponding to the tetraploid wheat chromosomes varied in length from 183.99 cM (3A) to 415.75 cM (2B), while the number of SNPs in each group varied from 57 (6A) to 158 (2A). A full list of the markers, including multiple SNPs merged into single loci, and their positions in all linkage groups is given in Table S2.

**Comparison between tetraploid and hexaploid wheat genetic maps**

A consensus genetic map of tetraploid wheat has previously been constructed using the iSelect 90k wheat SNP array (Maccaferri et al. 2015); some of the markers included on that array correspond to a subset of the Axiom SNPs, and were used to compare the 2 genetic maps. Of a total of 241 markers common to both maps, 237 were assigned to the same chromosome on both maps, and for these the marker order along each linkage group was also compared (Figures S2 & S3). In spite of the relatively small number of shared markers, there was clear co-linearity between the maps, as shown by the Spearman’s rank correlations between each pair of linkage groups (Table 2). Most of the chromosomes had correlation co-efficients >0.9; the lowest correlation of 0.787 (2A) is explained by the presence of a large group of markers with the same genetic position at 107-108 cM on the *T. durum* consensus map, which are resolved into several separate groups in our population. The genetic map was also compared with the consensus genetic map for hexaploid bread wheat previously constructed using the Axiom 820k wheat SNP array (Winfield et al. 2016), as it includes a much larger number of shared markers (808). Of these, 794 (98.3%) were assigned to the same chromosome on both maps; of those that were not, 8/14 were previously mapped to the D genome, but here were in the corresponding A or B linkage group, suggesting that they detect homoeologous loci.

Generally a high degree of co-linearity was observed between all 3 maps, but with some differences, exemplified by the maps for chromosomes 4B and 6B (Fig. 3). Often groups of markers that were co-located on the consensus maps (e.g. the large group at 48.9 cM in the *T. aestivum* 4B map, and that at 70.2 cM in the *T. aestivum* 6B map) were resolved to multiple loci in this study, meaning that our linkage groups were generally longer than the consensus maps. In other cases, groups of linked markers had a different location in one of the maps. For example, the segment from ~200 – 210 cM of chromosome 4B in our map was co-linear with a shared marker in the *T. durum* consensus map, but was placed at the proximal end of the same chromosome in the bread wheat map (Fig. 3A). This could indicate a putative intra-chromosomal re-arrangement in tetraploid wheat compared with bread wheat. Finally, the order of some adjacent markers on our map was reversed in compared to one of the consensus maps (e.g. AX-94517352 and AX-94685096 on 6B). Similar variations were also observed in some other chromosomes, and the full set of comparative maps is given in Figure S2-S4. Between 29-81 markers were shared between each linkage group and the hexaploid wheat map, depending on the chromosome (Table 2).

. The use of SNP markers meant that these could also be mapped to the publically available chromosome-specific *T. aestivum* var. Chinese Spring sequences (International Wheat Genome Sequencing Consortium 2014), by alignment of the SNP flanking sequences to the bread wheat sequence contigs. The majority of the SNPs (1186/1345, 88.2%) could be mapped to a sequence contig from the same chromosome, with 2A giving the most matches (151) and 4A the fewest (50). These marker locations will enable genetic map locations to be related to the bread wheat genome sequence, which will be of great value in identifying candidate genes for traits of interest.

**Changes in root morphology of F6 hybrid seedlings under osmotic stress**

As a preliminary screen for drought tolerance characteristics, seeds of all F6 hybrid lines were germinated on plant agar and allowed to develop in the dark.As described in Materials & Methods, agar plates were first infused either with medium only or with a solution of PEG-6000 to create osmotic stress, and imaged and analyzed 4 and 6 days after sowing (DAS); results are summarized in Figure 4. In most lines the osmotic stress resulted in a significant inhibition of root development (e.g. Fig. 4A); however, there were also some lines where the roots developed at a similar rate under osmotic stress, although shoot development still restricted (e.g. Fig. 4B). These lines demonstrated the ability rapidly to redirect metabolic effort to root development under osmotic stress. From the plant images, root length, surface area, average diameter, volume, and number of roots and forks were calculated for each seedling,and the population-wide variation for each trait was evaluated by 2-way ANOVA (Table 3). No significant variation in the number or average diameter of roots was observed under water stress (data not shown). However, across the whole population at 4 DAS, the osmotic stress treatment resulted in a 38-40% reduction in mean total root length (Fig. 4C), root volume (Fig. 4D), and surface area. As expected the statistical analysis indicated a highly significant difference between control and PEG-treated samples. While significant differences were not observed between the genotypes, there was a significant Genotype x Treatment (G x T) interaction, indicating that different genotypes responded differently to the osmotic stress. At 6 DAS the differences were slightly lower in magnitude depending on the trait measured, and all traits showed significant differences depending on genotype (p < 0.001), but only total root length showed a significant G x T interaction. In order to obtain a value that combined these data, the mean of root morphological measurements under osmotic stress was divided by that under control conditions for each line to give a stress ratio for root length (Fig. 4E), root volume (Fig. 4F), and surface area. The population distribution clustered around the average reduction in development noted above, but for each measurement a minority of lines had a stress ratio ≥ 1, indicating that these lines produced roots with equal or better efficiency during water deficit as in normal conditions, suggesting that they contained genetic loci that could confer drought tolerance.

**Association mapping of drought tolerance traits**

In order to assess the usefulness of the hybrid tetraploid wheat genetic map produced in this study, and identify putative QTLs and markers for drought tolerance traits, association mapping was carried out for the phenotypic trait data from the preliminary yield trials (Table S1) and root morphology analysis described above. In the absence of a complete tetraploid wheat genome sequence, the order and genetic positions of the markers in the genetic map were used as a surrogate for physical marker positions in the mapping procedure.

Underlying population structure was assessed by principal component analysis (PCA) including all the parental and hybrid lines. The first principal component accounted for 34.5% of the variation in the population and separated the lines by species. The parental lines from emmer wheat and durum wheat formed two tight, widely separated clusters (Fig. 5A). The F6 hybrids were distributed between these two clusters, but closer to the *T. durum* cluster, as expected from the initial backcross carried out when the population was established. After removing the parental lines, the 2nd and 3rd principal components (explaining 5.1% and 4.4% of the total variation) were plotted for all the F6 hybrids. As shown in Fig. 5B, the hybrids were loosely clustered on the basis of the parental *T. durum* genotype, but these clusters were also largely intermingled, reflecting the fact that all these genotypes originated from Turkish breeding programs. All subsequent principal components had relatively minor contributions; therefore, the first 3 components only were used as covariates in the association mapping calculations.

The SNPs that gave significant associations with phenotypic measurements are summarized in Table 4. From the preliminary yield trial data, there was an association between coleoptile length and a cluster of SNPs from chromosome 4B (Fig. 5C). The same SNPs were also associated with final plant height and lodging (data not shown). A second cluster with a slightly weaker assocation with coleoptile length and plant height was located on chr. 1A, but in this case had no association with increased lodging. A cluster of SNPs at the distal end of chromosome 6B was associated with anthocyanin accumulation (Fig. 5D).

From the seedling root morphology experiments, SNP clusters were identified that were associated with sustained seedling root development under osmotic stress. The most consistent of these was located on chromosome 1B and significantly associated with total root length (Fig. 5E), surface area, and root volume after 3 days of stress treatment. A second cluster on 6B was found to be associated with root volume but neither of the other traits, and only after 5 days of stress treatment (Fig. 5F), suggesting that it relates to a separate mechanism that has an effect after more prolonged osmotic stress. A third cluster of SNPs located on chromosome 5A was consistently observed to be weakly associated with root length and volume maintenance but did not cross the p = 0.001 significance threshold (Fig. 5E & F, orange ellipses). A larger study should be carried out to determine whether this represents a genuine marker-trait association.

The cM ranges of the SNP clusters associated with each trait were used to determine putative QTL locations on the genetic map, which are marked on Fig. 3. The combinations of SNP alleles corresponding to improved values for each trait were identified from the mapping statistics and are indicated in Table 4. For most of the markers, the most positive value for each trait was associated with the haplotype matching the majority of wild tetraploid parents, while the least positive value was given by the cultivated *T. durum* parental haplotype. The only exception was *Cln-q1,* where the *T. durum* parental haplotype was preferable.

**Discussion**

**Evaluation of the Wheat Breeders Array for use with tetraploid wheats**

Of the SNP arrays that are now commercially available for wheat genotyping, the Axiom 384HT Wheat Breeders’ Array offers substantially higher throughput, and therefore lower cost per sample, than previous platforms (Table 5). However, it contains a smaller number of SNPs (35,143) than other arrays, and these SNPs were selected primarily with breeding elite bread wheat genotypes in mind. The Wheat Breeders Array has been characterized for for a wide variety of hexaploid wheats (Allen et al. 2017). In this study, we wanted to ascertain whether it could be a cost-effective alternative for durum wheat breeding.

The SNP call rate for the durum wheat and F6 hybrid samples was similar to that observed for hexaploids, although the quality control cut-offs did have to be relaxed to include the parental emmer wheat lines. This was unexpected, as 11,360 of the SNPs had been mapped to the bread wheat D genome, and so were not expected to give meaningful data in tetraploid wheat. As the SNP chromosome assignments were made by aligning their flanking sequences to the best matching IWGSC contig (The International Wheat Genome Sequencing Consortium 2014; Winfield et al. 2016), it is possible that some match close homoeologs in addition to their reported location. During subsequent filtering, a high proportion of the ‘D genome’ SNPs failed quality criteria or were monomorphic, suggesting that homoeologous sequences from the A & B genomes were hybridizing to these probes but contained no genetic variation. Even so, 2054 SNPs assigned to the D genome were also polymorphic in this study. Similarly, a previous study using a set of 5386 hexaploid wheat SNPs also found that a minority of SNPs mapping to the D genome could also be genotyped in tetraploid wheat (Oliveira et al. 2014) Interestingly, of SNPs that were included in both our genetic map and the hexaploid wheat consensus map (Winfield et al. 2016), only 16 were previously mapped to the D genome.

Similarities between homoeologous chromosomes produce particular problems for SNP genotyping in polyploids, as they produce background signals that can confound automated SNP calling algorithms. This issue was expected to be exacurbated by the genetic variability present in our population, so the clustering results for all of the 9,484 polymorphic SNPs were visually inspected. Some genotype calls were manually corrected for 248 SNPs, 2.6% of the total, showing that the SNP calling algorithm was generally accurate but that some manual checking is necessary, as has also been reported for the Illumina iSelect array (Wang et al. 2014a). The majority of changed calls were for OTVs that had been missed by the calling algorithm.

Within our tetraploid population 9,484 array SNPs (27.0%) were polymorphic between at least 2 lines; this is a higher proportion than on the Axiom HD array (59,079/819,471) and similar to the proportion of iSelect array SNPs (26,626/91,829) that have been successfully mapped in durum wheat (Maccaferri et al. 2015; Winfield et al. 2016). However, some of the SNPs reported here were polymorphic only in a minority of the hybridizations tested; the number that were polymorphic in any individual cross varied from 2,930-5,894 depending on the parents. Between *T. durum* cultivars, the polymorphism rate would be lower, especially if they were closely related. In summary, the Wheat Breeders Array is a valuable genotyping alternative for tetraploid wheat, offering higher throughput and lower cost than other available wheat SNP arrays, provided that the parents of the population(s) to be genotyped are first tested to ensure that they include sufficient genetic variation.

**A genetic mapping strategy for pre-existing breeding lines**

Genetic mapping is a key technique for crop breeding, and high-throughput genotyping methods enable the production of genetic maps of very high marker density. This both increases the power of linkage mapping (Ott, Wang, and Leal 2015) and enables genome-wide association studies even in large-genome crops such as wheat (Bajgain et al. 2015). Recent studies have suggested that combining linkage and association mapping techniques offers improved QTL prediction in e.g. maize (Giraud et al. 2014; Ogut et al. 2015). However, most of these studies use multiple large biparental populations, which require substantial investment to produce. Therefore, there is a need for an approach which can be used effectively when resources are limited, or to characterize material from pre-existing breeding programmes. In this study we used a relatively simple method to combine linkage data from a relatively small population of inter-related families. It was assumed that family-specific effects would have a minimal impact on the recombination frequency, which seemed reasonable as all of the lines were propagated in the same conditions. The final map contained 1345 markers distributed across all 14 tetraploid wheat chromosomes (Figure 2), similar in density to the previous SSR/DArT marker framework map of *T. durum* (Maccaferri et al. 2014). Another advantage of this approach was that the map could readily be compared with consensus wheat maps generated using some of the same SNPs (Figure 3, Figures S2-S4). In general, the consistency of chromosome assignments and marker groupings between these datasets lends confidence to the accuracy of our map; at the same time, the consensus maps were required to resolve some ambiguities in the map order. We also identified a small number of markers that had different location in our map compared with one or both of the consensus maps (e.g. Fig. 3A, full list in Table S3). In these cases it we may have detected some chromosome rearrangements specific to tetraploid wheat, or to the wild emmer wheat lines used here. Previous cytogenetic studies have indicated that major chromosome rearrangements are relatively common between polyploid wheat species, and particularly in wild emmer wheat (Badaeva et al. 2007). It is important that breeders working with genetic material that is relatively distant from elite cultivars are aware of the existence of such translocations, again indicating the value of a genotyping study of this scale.

Our genetic map was sufficient to identify some marker-trait associations, although as some of the intervals between markers were large, there may be additional loci that could be identified with denser marker coverage. This could be achieved using much larger mapping populations from each cross, as in the high-density consensus tetraploid wheat SNP map (Maccaferri et al. 2015), in which case it is expected that a much larger proportion of the 9484 polymorphic SNPs described here would be incorporated into the final map.

**QTLs for abiotic stress tolerance in Turkish durum wheat**

In this study we identified SNP markers for 6 QTLs related to plant growth and osmotic stress tolerance in the F6 hybrid population (Table 4). Coleoptile length is related to final plant height, and the QTL identified on chromosome 4B was associated with both traits. The most significant genes contributing to control of these traits in modern wheat cultivars are *Rht-B1* and *Rht-D1*, mutations of which confer the semi-dwarf phenotype that drove the ‘green revolution’ (Peng et al. 1999). *Rht-B1* was previously mapped to chromosome 4BS (Ellis et al. 2002), and so was the likely candidate for the QTL observed here. The 4 SNPs identified here were mapped to the IWGSC sequence contigs (The International Wheat Genome Sequencing Consortium 2014), and indeed located to a 2 cM region on the 4BS Genome zipper that also contains *Rht-B1.*  The closest SNP marker in the virtual gene order, AX-95177742, also gave the strongest association with coleoptile length (p = 0.00011) and plant height (p = 0.00099).Most of the emmer wheat lines used in this study had longer coleoptiles than the durum cultivars, consistent with them containing the wild-type *Rht-B1* allele, while the durum wheat cultivars would be expected to contain the semi-dwarf mutant allele. Taken together these observations indicate that the coleoptile length/plant height QTL observed here is caused by the *Rht-B1* gene. Semi-dwarfism is desirable as it allows increased grain yield without the plant falling over under its own weight (‘lodging’). Accordingly, the wild-type *Rht-B1* allele was also associated with increased lodging in this study (p = 0.00026). While this is therefore not a useful target for breeding, it does confirm the effectiveness of our approach.

While the semi-dwarf trait is crucial for increasing wheat yield in optimal growth conditions, the short coleoptile length can make it less suitable for dry regions where deep sowing is desirable to utilize available soil moisture (Schillinger et al. 1998). Therefore, recent studies have identified a number of other QTLs for increasing coleoptile length in semi-dwarf wheat (Rebetzke et al. 2007; Li et al. 2016). The *Cln-q1* QTL on chr 1A associated with coleoptile length and plant height (Fig. 5C) is intriguing in this respect, as it was not associated with increased lodging; however, in this instance it seems that the cultivated durum wheats already had the optimum genotype among the lines tested here. Nevertheless, the wild emmer lines used in this study could be useful resource for mining for *Rht-B1* alleles that confer a more moderate phenotype than the *Rht-B1b* allele found in most commercial wheats, and the diversity of haplotypes observed for this locus in our population (Table 4) suggest that it would be effective for mapping this trait.

The early heading QTL indicated on chromosome 4B could also provide a yield advantage by allowing grain filling to begin earlier in the year, before summer droughts occur. A similar QTL was recently identified on chr 4B in a multiparental durum wheat RIL population (Milner et al. 2016), but does not appear to co-localize with our QTL, which was at the distal end of the chromosome. Two other early heading QTLs originating from wild emmer wheat have been identified in *T. aestivum* chromosome arm substitution lines, on chr 3AL & 7BS (Zhou et al. 2016). The *D2h-q1* QTL identified here warrants further investigation for its possible utility in durum wheat breeding.

Anthocyanins are natural anti-oxidants produced by some wheat lines, the accumulation of which is readily observable through red-purple colour in the stem and/or grain (Žofajová et al. 2012). Increasing the anthocyanin content of wheat can help to protect the plant from the damaging effects of reactive oxygen species produced during drought stress (Budak, Kantar, and Yucebilgili Kurtoglu 2013). Several of the wild emmer parents used in this study had a red colouration, and the *Anth-q1* QTL identified on chr 6B is a starting point for breeding for this useful trait.

Root system architecture is attracting increasing attention as a basis for optimizing yield under limited water conditions, including in durum wheat (Canè et al. 2014). A variety of architectural traits play an important role in drought avoidance (Maccaferri et al. 2016), and young seedling root development can be a useful indicator of the performance of adult plants (Placido et al. 2013). In this study, we identified 2 QTLs related to improved seedling root development under osmotic stress, located on chromosomes 1B and 6B, as well as a SNP cluster on 5A that was weakly associated, but not statistically significant (Fig. 5E-F). A new QTL for the response to exogenous abscisic acid (ABA) of seedling root length, shoot length and root growth inhibition was recently also reported in this latter region in hexaploid wheat (Iehisa et al. 2014). As ABA signalling is one of the major pathways controlling the response to osmotic stress in plants (Song et al. 2016), it would be interesting to investigate whether this association becomes significant under ABA treatment. Meanwhile, in a very detailed study of durum seedling root architecture combining data from both linkage mapping and an assocation panel, multiple QTLs were detected on all chromosomes (Maccaferri et al. 2016). In particular, the *RosS-q1* and *Rvol-q2* QTLs described here would fall within major root architecture QTL clusters identified on chr 1BS and 6BL respectively in the aforementioned study.

For all of the QTLs for traits related to drought tolerance or avoidance, the most advantageous haplotypes were found in the wild emmer wheat parents, as expected. The SNPs delineating these QTLs both serve as a guide for fine mapping or mining of wheat genome data to identify candidate genes, and provide molecular markers that will facilitate deployment of these traits in future breeding programs.

In conclusion, we demonstrated that the Axiom Wheat Breeder’s Genotyping array provides a credible and cost-effective alternative for high-throughput genotyping in tetraploid wheat. We were able to produce a densely populated genetic map and identify some useful marker-trait associations, although due to the limited size of the test population, the QTLs detected were those that have relatively large effects, such as the *Rht-B1* locus. It is expected that genotyping a larger F6 population would both increase the density of the genetic map and the power to detect QTLs. The loci identified in this study provide a useful starting point for ongoing breeding to develop durum wheat cultivars adapted to Turkish growing conditions.

**Acknowledgments**

This work was supported by the 1003 Primary Subjects R&D Funding Program of TÜBİTAK (The Scientific and Technological Research Council of Turkey, Grant no: 113O116) and by the General Directorate of Agriculture Research and Policy, Republic of Turkey Ministry of Food, Agriculture and Livestock (Grant no: TAGEM/TA/12/03/01/001). The authors thank İpek Özdemir for technical assistance with DNA isolation. We also thank the Wheat Improvement Strategic Programme (WISP) for designing the SNPs used on the Axiom array and making the marker information publicly available.

**References**

Allen, Alexandra M., Mark O. Winfield, Amanda J. Burridge, Rowena C. Downie, Harriet R. Benbow, Gary L. A. Barker, Paul A. Wilkinson, et al. 2017. “Characterization of a Wheat Breeders’ Array Suitable for High-Throughput SNP Genotyping of Global Accessions of Hexaploid Bread Wheat (*Triticum Aestivum*).” *Plant Biotechnology Journal* 15 (3): 390–401. doi:10.1111/pbi.12635.

Allen, Alexandra M, Gary L A Barker, Paul Wilkinson, Amanda Burridge, Mark Winfield, Jane Coghill, Cristobal Uauy, et al. 2012. “Discovery and Development of Exome-Based , Co-Dominant Single Nucleotide Polymorphism Markers in Hexaploid Wheat (Triticum aestivum L .).” *Plant Biotechnology Journal*, 1–17. doi:10.1111/pbi.12009.

Araus, J L, D Villegas, N Aparicio, L F G del Moral, S El Hani, Y Rharrabti, J P Ferrio, and C Royo. 2003. “Environmental Factors Determining Carbon Isotope Discrimination and Yield in Durum Wheat under Mediterranean Conditions.” *Crop Science* 43 (1): 170–80.

Badaeva, E D, O S Dedkova, G Gay, V a Pukhalskyi, a V Zelenin, S Bernard, and M Bernard. 2007. “Chromosomal Rearrangements in Wheat: Their Types and Distribution.” *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada* 50 (10): 907–26. doi:10.1139/G07-072.

Bajgain, P, M N Rouse, P Bulli, S Bhavani, T Gordon, R Wanyera, P N Njau, W Legesse, J A Anderson, and M O Pumphrey. 2015. “Association Mapping of North American Spring Wheat Breeding Germplasm Reveals Loci Conferring Resistance to Ug99 and Other African Stem Rust Races.” *BMC Plant Biology* 15: 249. doi:10.1186/s12870-015-0628-9.

Bérard, Aurélie, Marie Christine Le Paslier, Mireille Dardevet, Florence Exbrayat-Vinson, Isabelle Bonnin, Alberto Cenci, Annabelle Haudry, Dominique Brunel, and Catherine Ravel. 2009. “High-Throughput Single Nucleotide Polymorphism Genotyping in Wheat (Triticum Spp.).” *Plant Biotechnology Journal* 7 (4): 364–74. doi:10.1111/j.1467-7652.2009.00404.x.

Bradbury, Peter J., Zhiwu Zhang, Dallas E. Kroon, Terry M. Casstevens, Yogesh Ramdoss, and Edward S. Buckler. 2007. “TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples.” *Bioinformatics* 23 (19): 2633–35. doi:10.1093/bioinformatics/btm308.

Brenchley, Rachel, Manuel Spannagl, Matthias Pfeifer, Gary L a Barker, Rosalinda D’Amore, Alexandra M Allen, Neil McKenzie, et al. 2012. “Analysis of the Bread Wheat Genome Using Whole-Genome Shotgun Sequencing.” *Nature* 491 (7426). Nature Publishing Group: 705–10. doi:10.1038/nature11650.

Budak, Hikmet, Melda Kantar, and Kuaybe Yucebilgili Kurtoglu. 2013. “Drought Tolerance in Modern and Wild Wheat.” *The Scientific World Journal*. doi:10.1155/2013/548246.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10: 421. doi:10.1186/1471-2105-10-421.

Canè, Maria Angela, Marco Maccaferri, Ghasemali Nazemi, Silvio Salvi, Rossella Francia, Chiara Colalongo, and Roberto Tuberosa. 2014. “Association Mapping for Root Architectural Traits in Durum Wheat Seedlings as Related to Agronomic Performance.” *Molecular Breeding* 34 (4): 1629–45. doi:10.1007/s11032-014-0177-1.

Challinor, A J, J Watson, D B Lobell, S M Howden, D R Smith, and N Chhetri. 2014. “A Meta-Analysis of Crop Yield under Climate Change and Adaptation.” *Nature Climate Change* 27 (March): 1–5. doi:10.1038/NCLIMATE2153.

de Givry, Simon, Martin Bouchez, Patrick Chabrier, Denis Milan, and Thomas Schiex. 2005. “CARTHAGENE: Multipopulation Integrated Genetic and Radiation Hybrid Mapping.” *Bioinformatics* 21 (8): 1703–4. doi:10.1093/bioinformatics/bti222.

Ellis, Marc H., W. Spielmeyer, K. R. Gale, G. J. Rebetzke, and R. A. Richards. 2002. “‘Perfect’ markers for the Rht-B1b and Rht-D1b Dwarfing Genes in Wheat.” *Theoretical and Applied Genetics* 105 (6–7): 1038–42. doi:10.1007/s00122-002-1048-4.

Endelman, Jeffrey B, Jean-Luc Jannink, and Robert W Holley. 2012. “Shrinkage Estimation of the Realized Relationship Matrix.” *G3:Genes, Genomes, Genetics* 2: 1405–1413. doi:doi: 10.1534/g3.112.004259.

Ergen, Neslihan Z, and Hikmet Budak. 2009. “Sequencing over 13 000 Expressed Sequence Tags from Six Subtractive cDNA Libraries of Wild and Modern Wheats Following Slow Drought Stress.” *Plant, Cell & Environment* 32 (3). Blackwell Publishing Ltd: 220–36. doi:10.1111/j.1365-3040.2008.01915.x.

Giraud, Héloïse, Christina Lehermeier, Eva Bauer, Matthieu Falque, Vincent Segura, Cyril Bauland, Christian Camisan, et al. 2014. “Linkage Disequilibrium with Linkage Analysis of Multiline Crosses Reveals Different Multiallelic QTL for Hybrid Performance in the Flint and Dent Heterotic Groups of Maize.” *Genetics* 198 (4): 1717–34. doi:10.1534/genetics.114.169367.

Giunta, F., R. Motzo, and M. Deidda. 1993. “Effect of Drought on Yield and Yield Components of Durum Wheat and Triticale in a Mediterranean Environment.” *Field Crops Research* 33 (4). Elsevier: 399–409. doi:10.1016/0378-4290(93)90161-F.

Holland, James B. 2015. “MAGIC Maize: A New Resource for Plant Genetics.” *Genome Biology* 16 (1): 163. doi:10.1186/s13059-015-0713-2.

Iehisa, Julio C M, Takakazu Matsuura, Izumi C. Mori, and Shigeo Takumi. 2014. “Identification of Quantitative Trait Locus for Abscisic Acid Responsiveness on Chromosome 5A and Association with Dehydration Tolerance in Common Wheat Seedlings.” *Journal of Plant Physiology* 171 (2): 25–34. doi:10.1016/j.jplph.2013.10.001.

International Grains Council. 2016. “Grain Market Report.” London. http://www.igc.int/downloads/gmrsummary/gmrsumme.pdf.

International Wheat Genome Sequencing Consortium. 2014. “A Chromosome-Based Draft Sequence of the Hexaploid Bread Wheat (Triticum Aestivum) Genome.” *Science (New York, N.Y.)* 345 (6194): 1250092. doi:10.1126/science.1251788.

Kosambi, D. D. 1943. “The Estimation of Map Distances From Recombination Values.” *Annals of Eugenics* 12: 172–75. doi:10.1111/j.1469-1809.1943.tb02321.x.

Li, Genqiao, Guihua Bai, Brett F. Carver, Norman C. Elliott, Rebecca S. Bennett, Yanqi Wu, Robert Hunger, J. Michael Bonman, and Xiangyang Xu. 2016. “Genome-Wide Association Study Reveals Genetic Architecture of Coleoptile Length in Wheat.” *Theoretical and Applied Genetics*, November. Springer Berlin Heidelberg, 1–11. doi:10.1007/s00122-016-2820-1.

Lorenc, Michał T., Satomi Hayashi, Jiri Stiller, Hong Lee, Sahana Manoli, Pradeep Ruperao, Paul Visendi, et al. 2012. “Discovery of Single Nucleotide Polymorphisms in Complex Genomes Using SGSautoSNP.” *Biology*. doi:10.3390/biology1020370.

Lorieux, Mathias. 2012. “MapDisto: Fast and Efficient Computation of Genetic Linkage Maps.” *Molecular Breeding* 30 (2): 1231–35. doi:10.1007/s11032-012-9706-y.

Lucas, Stuart J, Bala Akpınar, Hana Šimková, Marie Kubaláková, Jaroslav Doležel, and Hikmet Budak. 2014. “Next-Generation Sequencing of Flow-Sorted Wheat Chromosome 5D Reveals Lineage-Specific Translocations and Widespread Gene Duplications.” *BMC Genomics* 15 (1): 1080. doi:10.1186/1471-2164-15-1080.

Maccaferri, Marco, Maria Angela Cane’, Maria C Sanguineti, Silvio Salvi, Maria C Colalongo, Andrea Massi, Fran Clarke, et al. 2014. “A Consensus Framework Map of Durum Wheat (Triticum Durum Desf.) Suitable for Linkage Disequilibrium Analysis and Genome-Wide Association Mapping.” *BMC Genomics* 15 (1): 873. doi:10.1186/1471-2164-15-873.

Maccaferri, Marco, Walid El-Feki, Ghasemali Nazemi, Silvio Salvi, Maria Angela Canè, Maria Chiara Colalongo, Sandra Stefanelli, and Roberto Tuberosa. 2016. “Prioritizing Quantitative Trait Loci for Root System Architecture in Tetraploid Wheat.” *Journal of Experimental Botany* 67 (4): 1161–78. doi:10.1093/jxb/erw039.

Maccaferri, Marco, Andrea Ricci, Silvio Salvi, Sara Giulia Milner, Enrico Noli, Pier Luigi Martelli, Rita Casadio, et al. 2015. “A High-Density, SNP-Based Consensus Map of Tetraploid Wheat as a Bridge to Integrate Durum and Bread Wheat Genomics and Breeding.” *Plant Biotechnology Journal* 13 (5): 648–63. doi:10.1111/pbi.12288.

McMullen, Michael D., Stephen Kresovich, Hector Sanchez Villeda, Peter Bradbury, Huihui Li, Qi Sun, Sherry A. Flint-Garcia, et al. 2009. “Genetic Properties of the Maize Nested Association Mapping Population.” *Science (New York, N.Y.)* 325 (5941): 737–40. doi:10.1126/science.1174320.

Milner, Sara Giulia, Marco Maccaferri, Bevan Emma Huang, Paola Mantovani, Andrea Massi, Elisabetta Frascaroli, Roberto Tuberosa, and Silvio Salvi. 2016. “A Multiparental Cross Population for Mapping QTL for Agronomic Traits in Durum Wheat ( *Triticum Turgidum* Ssp. *Durum* ).” *Plant Biotechnology Journal* 14 (2): 735–48. doi:10.1111/pbi.12424.

Ogut, F, Y Bian, P J Bradbury, and J B Holland. 2015. “Joint-Multiple Family Linkage Analysis Predicts within-Family Variation Better than Single-Family Analysis of the Maize Nested Association Mapping Population.” *Heredity* 114 (6): 552–63. doi:10.1038/hdy.2014.123.

Oliveira, Hugo R, Jenny Hagenblad, Matti W Leino, Fiona J Leigh, Diane L Lister, Leonor Penã-Chocarro, and Martin K Jones. 2014. “Wheat in the Mediterranean Revisited--Tetraploid Wheat Landraces Assessed with Elite Bread Wheat Single Nucleotide Polymorphism Markers.” *BMC Genetics* 15: 54. doi:10.1186/1471-2156-15-54.

Ott, Jurg, Jing Wang, and Suzanne M Leal. 2015. “Genetic Linkage Analysis in the Age of Whole-Genome Sequencing.” *Nature Reviews Genetics* 16 (5): 275–84. doi:10.1038/nrg3908.

Peng, J, D E Richards, N M Hartley, and G P Murphy. 1999. “‘Green Revolution’genes Encode Mutant Gibberellin Response Modulators.” *“Green Revolution”genes Encode Mutant Gibberellin Response Modulators* 400 (July): 8–13. doi:10.1038/22307.

Placido, Dante F, Malachy T Campbell, Jing J Folsom, Xinping Cui, Greg R Kruger, P Stephen Baenziger, and Harkamal Walia. 2013. “Introgression of Novel Traits from a Wild Wheat Relative Improves Drought Adaptation in Wheat.” *Plant Physiology* 161 (4): 1806–19. doi:10.1104/pp.113.214262.

Rebetzke, G. J., M. H. Ellis, D. G. Bonnett, and R. A. Richards. 2007. “Molecular Mapping of Genes for Coleoptile Growth in Bread Wheat (Triticum Aestivum L.).” *Theoretical and Applied Genetics* 114 (7): 1173–83. doi:10.1007/s00122-007-0509-1.

Rice, William E. R. 1989. “Analyzing Tables of Statistical Tests.” *Evolution* 43: 223–25. doi:10.2307/2409177.

Schillinger, William F., Edwin Donaldson, Robert E. Allan, and Stephen S. Jones. 1998. “Winter Wheat Seedling Emergence from Deep Sowing Depths.” *Agronomy Journal* 90 (5): 582–86. doi:10.2134/agronj1998.00021962009000050002x.

Song, L., S.-s. C. Huang, A. Wise, R. Castanon, J. R. Nery, H. Chen, M. Watanabe, J. Thomas, Z. Bar-Joseph, and J. R. Ecker. 2016. “A Transcription Factor Hierarchy Defines an Environmental Stress Response Network.” *Science* 354 (6312): aag1550-aag1550. doi:10.1126/science.aag1550.

Tanksley, Steven D., and Susan R. McCouch. 1997. “Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild.” *Science* 277 (5329).

Tardieu, Franois. 2012. “Any Trait or Trait-Related Allele Can Confer Drought Tolerance: Just Design the Right Drought Scenario.” *Journal of Experimental Botany* 63 (1): 25–31. doi:10.1093/jxb/err269.

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 (6194): 1251788. doi:10.1126/science.1251788.

Trebbi, Daniele, Marco Maccaferri, Peter de Heer, Anker Sørensen, Silvia Giuliani, Silvio Salvi, Maria Corinna Sanguineti, Andrea Massi, Edwin Andries Gerard van der Vossen, and Roberto Tuberosa. 2011. “High-Throughput SNP Discovery and Genotyping in Durum Wheat (Triticum Durum Desf.).” *Theoretical and Applied Genetics* 123: 555–69. doi:10.1007/s00122-011-1607-7.

van Poecke, Remco M P, Marco Maccaferri, Jifeng Tang, Hoa T. Truong, Antoine Janssen, Nathalie J. van Orsouw, Silvio Salvi, Maria C. Sanguineti, Roberto Tuberosa, and Edwin A G van der Vossen. 2013. “Sequence-Based SNP Genotyping in Durum Wheat.” *Plant Biotechnology Journal* 11 (7): 809–17. doi:10.1111/pbi.12072.

Verslues, Paul E., Manu Agarwal, Surekha Katiyar-Agarwal, Jianhua Zhu, and Jian Kang Zhu. 2006. “Methods and Concepts in Quantifying Resistance to Drought, Salt and Freezing, Abiotic Stresses That Affect Plant Water Status.” *Plant Journal* 45 (4): 523–39. doi:10.1111/j.1365-313X.2005.02593.x.

Voorrips, R E. 2002. “MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs.” *The Journal of Heredity* 93 (1): 77–78. doi:10.1093/jhered/93.1.77.

Wang, Shichen, Debbie Wong, Kerrie Forrest, Alexandra Allen, Shiaoman Chao, Bevan E. Huang, Marco Maccaferri, et al. 2014a. “Characterization of Polyploid Wheat Genomic Diversity Using a High-Density 90 000 Single Nucleotide Polymorphism Array.” *Plant Biotechnology Journal* 12 (6): 787–96. doi:10.1111/pbi.12183.

Wang, Shichen, Debbie Wong, Kerrie Forrest, Alexandra Allen, Shiaoman Chao, Bevan E Huang, Marco Maccaferri, et al. 2014b. “Characterization of Polyploid Wheat Genomic Diversity Using a High-Density 90 000 Single Nucleotide Polymorphism Array.” *Plant Biotechnology Journal*, March, 1–10. doi:10.1111/pbi.12183.

Winfield, Mark O., Alexandra M. Allen, Amanda J. Burridge, Gary L A Barker, Harriet R. Benbow, Paul A. Wilkinson, Jane Coghill, et al. 2016. “High-Density SNP Genotyping Array for Hexaploid Wheat and Its Secondary and Tertiary Gene Pool.” *Plant Biotechnology Journal* 14 (5): 1195–1206. doi:10.1111/pbi.12485.

Zhou, Wei, Shasha Wu, Mingquan Ding, Jingjuan Li, Zhaobin Shi, Wei Wei, Jialian Guo, Hua Zhang, Yurong Jiang, and Junkang Rong. 2016. “Mapping of Ppd-B1, a Major Candidate Gene for Late Heading on Wild Emmer Chromosome Arm 2BS and Assessment of Its Interactions with Early Heading QTLs on 3AL.” *PloS One* 11 (2). Public Library of Science: e0147377. doi:10.1371/journal.pone.0147377.

Žofajová, Alžbeta, Ivana Pšenáková, Michaela Havrlentová, and Michaela Piliarová. 2012. “Accumulation of Total Anthocyanins in Wheat Grain.” *Agriculture* 58 (2): 50–56. doi:10.2478/v10207-012-0006-7.

**Table 1.** Classification of genotype calls for SNPs from each sub-genome in tetraploid wheats. Sub-genome locations were predicted by aligning SNP flanking sequences to the IWGSC chromosome-specific sequence contigs (International Wheat Genome Sequencing Consortium 2014).

|  |  |  |  |
| --- | --- | --- | --- |
| **Sub-genome** | **AA** | **BB** | **DD** |
| Total SNP no. | 10819 (100%) | 12932 (100%) | 11360 (100%) |
| Polymorphic, high resolution | 3268 (30.2%) | 3783 (29.3%) | 2054 (18.1%) |
| Monomorphic, high resolution | 3949 (36.5%) | 4742 (36.7%) | 5634 (49.6%) |
| No minor homozygote allele | 1077 (9.9%) | 1338 (10.3%) | 1159 (10.2%) |
| Off Target Variant | 158 (1.46%) | 247 (1.91%) | 94 (0.83%) |
| Below quality thresholds | 2367 (21.9%) | 2822 (21.8%) | 2419 (21.3%) |

**Table 2.** Marker statistics for tetraploid wheat hybrid genetic map and comparisons with previous genetic maps.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Linkage Group** | **Hybrid tetraploid** | | | **Markers shared with other wheat datasets** | | |
|  | SNPs | Genetic loci | Length (cM) | *T. durum*1 | *T. aestivum*2 | IWGSC3 |
| 1A | 98 | 69 | 380.91 | 15 (0.944) | 62 | 89 |
| 1B | 121 | 83 | 316.07 | 23 (0.853) | 81 | 105 |
| 2A | 158 | 78 | 223.47 | 34 (0.787) | 60 | 151 |
| 2B | 128 | 84 | 415.75 | 19 (0.905) | 78 | 116 |
| 3A | 79 | 49 | 183.99 | 15 (0.963) | 49 | 73 |
| 3B | 108 | 77 | 275.20 | 12 (0.949) | 60 | 97 |
| 4A | 61 | 54 | 249.72 | 8 (0.928) | 43 | 50 |
| 4B | 70 | 54 | 247.82 | 17 (0.821) | 37 | 62 |
| 5A | 96 | 60 | 261.06 | 20 (0.892) | 71 | 75 |
| 5B | 60 | 50 | 297.81 | 9 (0.932) | 29 | 55 |
| 6A | 57 | 36 | 213.39 | 10 (0.864) | 49 | 52 |
| 6B | 76 | 65 | 234.83 | 10 (0.966) | 55 | 57 |
| 7A | 147 | 103 | 406.76 | 32 (0.991) | 61 | 131 |
| 7B | 86 | 69 | 264.80 | 13 (0.972) | 59 | 73 |
| **Total** | **1345** | **931** | **3971.6** | **237** | **794** | **1186** |

1High-density durum wheat consensus map, Maccaferri *et al.* (2015). Data is given as no. of shared markers (Spearman’s rank correlation coefficient)

2High-density bread wheat consensus map, Winfield *et al.* (2016)

3Number of SNPs mapped to *T. aestivum* cv. Chinese Spring chromosome-specific survey sequence contigs (International Wheat Genome Sequencing Consortium 2014)

**Table 3.** Means and summary statistics for root morphology traits in response to osmotic stress treatment for 99 hybrid tetraploid wheat lines at 4 and 6 DAS. p-values were calculated using 2-way ANOVA with the random model for effects of genotype and treatment.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measurement | 4 days after sowing | | | 6 days after sowing | | |
| Trait (units) | Total root length (cm) | Surface area (cm.2) | Volume  (cm.3) | Total root length (cm) | Surface area (cm.2) | Volume  (cm.3) |
| Min | 0.59 | 0.079 | 0.001 | 0.32 | 0.10 | 0.001 |
| Max | 29.98 | 6.77 | 0.204 | 39.4 | 9.64 | 0.258 |
| Mean ± SD (control) | 15.7 ± 4.9 | 3.3 ± 1.1 | 0.060 ± 0.027 | 20.1 ± 6.9 | 4.5 ± 1.6 | 0.089 ± 0.047 |
| Mean ± SD (PEG) | 9.8 ± 4.6 | 2.0 ± 0.98 | 0.037 ± 0.025 | 15.8 ± 6.3 | 3.2 ± 1.3 | 0.059 ± 0.036 |
| p [Genotype] | 0.130 | 0.182 | 0.043 | 0.0038 | 0.0039 | 0.00047 |
| p [Treatment] | < 0.00001 | < 0.00001 | < 0.00001 | < 0.00001 | < 0.00001 | < 0.00001 |
| p [G x T] | < 0.00001 | < 0.00001 | < 0.00001 | 0.000056 | 0.108 | 0.085 |

**Table 4.** SNPs delineating putative QTLs for traits related to drought tolerance, identified by association mapping. The ‘Haplotypes’ column lists all combinations of SNP(s) observed in the population, with the haplotype making the largest positive contribution to the trait of interest given first, followed by other observed combinations in descending order.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Locus | Trait(s) | Associated SNPs | Chr. | Map position/  range (cM) | Haplotypes |
| *Rht-B1* | Coleoptile length, plant height, lodging | AX-94517352  AX-94525193  AX-94685096  AX-95177742 | 4B | 66.48 - 71.12 | CC:AA:CC:AA  CC:AA:TT:AA  TT:GG:CC:AA  CC:AA:TT:GG  TT:GG:TT:AA  TT:GG:CC:GG  TT:GG:TT:GG |
| *Cln-q1* | Coleoptile length, plant height | AX-94423107  AX-94782613  AX-94631437  AX-95226312  AX-95011371  AX-94381913  AX-95186496 | 1A | 79.82 – 82.85 | TT:AA:TT:AA:TT:CC:GG  TT:AA:CC:CC:CC:TT:AA  CC:CC:TT:AA:TT:CC:GG  CC:CC:CC:AA:TT:CC:AA  CC:CC:CC:CC:CC:TT:AA |
| *D2h-q1* | Days to heading | AX-94494832  AX-94605238  AX-94751091  AX-94575655 | 4B | 244.05 - 247.82 | CC:CC:GG:CC  CC:CC:GG:TT  CC:CC:CC:TT  GG:GG:CC:TT |
| *RosS-q1* | Root osmotic stress ratio | AX-94550967 AX-94980178 AX-95253262  AX-94574509 | 1B | 0 – 15.5 | CC:GG:GG:TT  CC:GG:CC:CC  TT:AA:GG:TT  TT:AA:CC:CC |
| *Rvol-q1* | Root volume stress ratio | AX-94916820  AX-94712929  AX-94492424  AX-94617650  AX-95143478  AX-95202867  AX-94671774 | 6B | 81.67-108.32 | GG:AA:CC:AA:GG:AA:GG  GG:AA:CC:AA:CC:AA:TT  GG:GG:CC:AA:GG:AA:GG  CC:AA:CC:AA:GG:AA:GG  CC:GG:CC:AA:GG:AA:GG  CC:AA:GG:AA:GG:AA:GG  CC:AA:GG:AA:AA:AA:GG  CC:GG:GG:AA:GG:AA:GG  CC:GG:GG:GG:AA:GG:TT |
| *Anth-q1* | Anthocyanin accumulation | AX-94881172  AX-94429674  AX-94823159  AX-94982539 | 6B | 168.90 – 173.07 | TT:GG:CC:GG  TT:GG:CC:AA  TT:AA:TT:GG  CC:GG:CC:AA  TT:AA:TT:AA  CC:AA:TT:GG  CC:AA:TT:AA |

**Table 5.** Comparison of currently available high throughput wheat genotyping arrays.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Axiom Wheat Breeder’s Genotyping Array | Axiom Wheat HD Genotyping Array | Infinium iSelect 90k Wheat SNP Array |
| Total no. of SNPs included | 35,143 | 819,571 | 91,829 |
| SNPs verified to be informative in diverse wheat accessions1 | 35,143 | 546,299 (66.7%) | 56,388 (61.4%) |
| SNPs verified to be informative in tetraploids1 | 9,484 (27%, this study) | 59,079 (7.2%) | 26,626 (28.9%) |
| Samples processed per array | 384 | 96 | 24 |
| Minimum samples/order | 768 | 96 | 48 |
| Per sample cost estimate2 | € 75 | € 450 | € 330 |

1Number of SNPs verified to be informative reported by Winfield *et al.*, 2015 (Axiom HD) and Wang *et al.,* 2014 (iSelect). The SNPs on the Wheat Breeder’s Array are a subset of those found to be informative on the Wheat HD array.

2Estimates based on quotations including arrays, consumeables and service costs for genotyping of 1,920 samples.

**Figure legends**

**Figure 1:** Clustering of genotype calls from tetraploid wheats. For all plots, marker shading indicates species; white = F6 hybrids, light gray = *T. dicoccon*, gray = *T. durum*, black = *T. dicoccoides*. The single *T. araraticum* sample is black and circled. **A**: Scatter plot for all samples passing quality filters for % call rate for all SNPs (y-axis) vs. % heterozygous calls (x-axis).  **B-F**: Scatter plots showing typical genotype clustering results for individual SNPs (ID above each plot). Marker shapes indicate the genotype call: ▲= Homozygous AA, ▼= Homozygous BB, ◌ = Heterozygous AB, = Off-target variant (OTV), □ = no call. **B.** Polymorphic SNP with symmetrical clustering. **C.** Polymorphic SNP with asymmetrical clustering due to tetraploid background.  **D.** ‘NoMinorHom’ SNP with ambiguous clusters. **E.** Correctly called OTV.  **F.** Mis-called OTV (OTV samples called as AA by the clustering algorithm).

**Figure 2:** Consensus linkage map for the hybrid tetraploid population for the A (top) and B (bottom) sub-genomes. Where multiple markers mapped to the same genetic position, only one is shown.

**Figure 3:** Comparison of SNP marker order in the linkage maps of chromosomes generated in this study (centre) with the *T. durum* (left) and *T. aestivum* (right) consensus genetic maps. Only markers shared between 2 or more of the 2 maps are shown. Chromosome positions are marked in centiMorgans (cM). **A.** Comparison for chromosome 4B. **B.** Comparison for chromosome 6B. Putative QTL locations determined by association mapping (Table 4) are marked with black bars.

**Figure 4:** Differences in root morphology of seedlings under osmotic stress or control conditions. **A-B.** Scans of typical seedlings at 4 DAS grown on plant agar infused with PEG-6000 to generate final water pressure of -0.7 Mpa (PEG) or with medium only (Control). **C, E.** Box and whisker plots showing quartile distribution of total root length (**C.**) and root volume (**E.**) measurements in all F6 hybrid lines under each condition. **D, F.** Histograms showing the ratio of measurements of total root length (**D.**) and root volume (**F.**) under osmotic stress divided by those under control conditions, for each individual line.

**Figure 5:** Association mapping of shoot and root traits related to drought tolerance. **A.** Principal component analysis (PCA) for all samples including parental lines. **B.** Plot of the second and third principal components for all F6 hybrids, coloured by *T. durum* parent. **C.-F.** Manhattan plots of marker-trait associations. Horizontal black lines indicate a significance threshold of p < 0.001, clusters of markers with significant associations are marked with blue ellipses. **C.** Marker-trait associations for coleoptile length. **D.** Marker-trait associations for anthocyanin accumulation. **E.** Marker-trait associations for root length stress ratio. **F.** Marker-trait assocations for root volume stress ratio after 5 days of treatment.

**Figure 1**

****

**Figure 2**

****

****

**Figure 3**

****

**Figure 4**

****

**Figure 5**



**Supporting Information**

**Figure S1.** QQ plot for marker association p-values with phenotypic traits measured in preliminary yield trials.

**Figure S2.** Genetic map comparisons of group 1 & 2 chromosomes from this study with previously published consensus maps for tetraploid and hexaploid wheat. Only markers with positions on at least 2 of the maps are shown.

**Figure S3.**  Genetic map comparisons of group 3 & 4 chromosomes from this study with previously published consensus maps for tetraploid and hexaploid wheat. Only markers with positions on at least 2 of the maps are shown.

**Figure S4.** Genetic map comparisons of group 5, 6 & 7 chromosomes from this study with previously published consensus maps for tetraploid and hexaploid wheat. Only markers with positions on at least 2 of the maps are shown.

**Table S1.** Parentage information of the 100 F6 hybrid lines used in the study, along with phenotypic data from the preliminary yield trials.

**Table S2.** Table of all markers included in the tetraploid wheat consensus genetic map including minor linkage groups, along with inter-marker genetic distances (cM).

**Table S3.** Table showing comparison of genetic positions of markers mapped in this study with their positions in previously published wheat consensus genetic maps.