



RESEARCH PAPER

Unraveling the genetic architecture of grain size in einkorn wheat through linkage and homology mapping and transcriptomic profiling

Kang Yu^{1,2,4,5,†}, Dongcheng Liu^{1,†}, Yong Chen^{6,†}, Dongzhi Wang^{1,3}, Wenlong Yang¹, Wei Yang⁴, Lixin Yin⁴, Chi Zhang^{4,5}, Shancen Zhao^{4,5}, Jiazhu Sun¹, Chunming Liu^{2,*}, and Aimin Zhang^{1,*}

¹ State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology/Innovation Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

² Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Beijing Genomics Institute-Shenzhen, Shenzhen 518083, China

⁵ State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China

⁶ Science and Technology Department, State Tobacco Monopoly Administration, Beijing 100045, China

† These authors contributed equally to this work.

* Correspondence: amzhang@genetics.ac.cn and cmliu@ibcas.ac.cn

Received 12 August 2018; Editorial decision 14 May 2019; Accepted 15 May 2019

Editor: Greg Rebetzke, CSIRO Agriculture and Food, Australia

Abstract

Understanding the genetic architecture of grain size is a prerequisite to manipulating grain development and improving the potential crop yield. In this study, we conducted a whole genome-wide quantitative trait locus (QTL) mapping of grain-size-related traits by constructing a high-density genetic map using 109 recombinant inbred lines of einkorn wheat. We explored the candidate genes underlying QTLs through homologous analysis and RNA sequencing. The high-density genetic map spanned 1873 cM and contained 9937 single nucleotide polymorphism markers assigned to 1551 bins on seven chromosomes. Strong collinearity and high genome coverage of this map were revealed by comparison with physical maps of wheat and barley. Six grain size-related traits were surveyed in five environments. In total, 42 QTLs were identified; these were assigned to 17 genomic regions on six chromosomes and accounted for 52.3–66.7% of the phenotypic variation. Thirty homologous genes involved in grain development were located in 12 regions. RNA sequencing identified 4959 genes differentially expressed between the two parental lines. Twenty differentially expressed genes involved in grain size development and starch biosynthesis were mapped to nine regions that contained 26 QTLs, indicating that the starch biosynthesis pathway plays a vital role in grain development in einkorn wheat. This study provides new insights into the genetic architecture of grain size in einkorn wheat; identification of the underlying genes enables understanding of grain development and wheat genetic improvement. Furthermore, the map facilitates quantitative trait mapping, map-based cloning, genome assembly, and comparative genomics in wheat taxa.

Keywords: Einkorn wheat (*Triticum monococcum*), grain size, high-density genetic map, quantitative trait loci, RAD-seq, RNA-seq.

Abbreviations:ADPG, ADP-glucose; ANOVA, analysis of variance; CI, confidence interval; CIM, composite interval mapping; cM, centimorgan; DAF, days after flowering; DEG, differentially expressed gene; FPKM, fragments per kilobase of transcript per million mapped reads; GA, grain area; GC, grain circumference; GL, grain length; GLW, grain length/width; GO, Gene Ontology; GW, grain width; InDel, insertion–deletion; KEGG, Kyoto Encyclopedia of Genes and Genomes; LOD, logarithm of the odds; QTL, quantitative trait locus; RAD-seq, restriction site-associated DNA sequencing; RIL, recombinant inbred line; RNA-seq, RNA sequencing; SNP, single nucleotide polymorphism; TGW, thousand-grain weight.

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Grain weight, which is underpinned by grain morphology, including two main components, grain length and width, is one of the most important traits in wheat (*Triticum aestivum* L.). During the domestication process and breeding history of wheat, grain size was a major selection and breeding target that has been widely selected and manipulated to increase grain yield (Gegas *et al.*, 2010). In China, improvement in wheat yield from $\sim 1 \text{ T ha}^{-1}$ in 1965 to $\sim 5.4 \text{ T ha}^{-1}$ today may be due to the large increase in grain weight (He *et al.*, 2001; Wu *et al.*, 2014). Grain morphology directly influences milling performance and seedling vigor, which in turn determines the end products (Campbell *et al.*, 1999; Gegas *et al.*, 2010). Large variations in grain size and weight occur among domesticated and wild species of diploid, tetraploid, and hexaploid wheat (Jing *et al.*, 2007; Gegas *et al.*, 2010). Thus, understanding the genetic factors underlying grain size would provide the prerequisite information necessary to improve wheat yields.

High-density genetic maps play a fundamental role in dissecting the genetic components of agronomic traits and assembling genomes. As a basic tool, high-density genetic maps have been widely developed for crop species, including cereal crops such as wheat (Iehisa *et al.*, 2014; Kumar *et al.*, 2016), rice (Xie *et al.*, 2010), maize (Chen *et al.*, 2014), and barley (Chutimanitsakun *et al.*, 2011), and economic crops such as eggplant (Barchi *et al.*, 2012), grape (Wang *et al.*, 2012a), and sesame (Zhang *et al.*, 2013). A high-density consensus map of tetraploid wheat was developed by integrating datasets of 13 biparental populations that harbored 30 144 markers and covered 2631 centimorgan (cM) of the A and B subgenomes (Maccaferri *et al.*, 2015). In hexaploid wheat, a high-density genetic map was recently constructed that included 119 566 single nucleotide polymorphism (SNP) markers, greatly facilitating the fine-mapping of a major quantitative trait locus (QTL) for grain number (Cui *et al.*, 2017). In barley, a high-density amplified fragment-length polymorphism map of 3H involving 84 markers covered 6.7 cM and was applied to narrow the genomic region for the important domestication loci, *Brittle rachis* (*Btr1* and *Btr2*), which have been molecularly cloned and characterized (Komatsuda *et al.*, 2004; Pourkheirandish *et al.*, 2015). Recently, with the demand for genome sequencing, high-density genetic maps have been widely exploited for genome assembly, especially to construct chromosomal pseudomolecules by anchoring and ordering scaffolds in wheat (Jia *et al.*, 2013; Chapman *et al.*, 2015), cotton (Zhang *et al.*, 2015), and peanut (Bertioli *et al.*, 2016).

For species with less available genomic information, there is an urgent need to develop a high-density genetic map with many genetic markers distributed over the entire genome. Advances in high-throughput sequencing technologies provide an excellent platform for genome-wide discovery of sequence variation and the development of polymorphic genetic markers. Of these technologies, genotyping-by-sequencing methods utilize restriction enzyme digestion to reduce the complexity of a genome and sequence large amounts of nucleotide fragments using next-generation sequencing (NGS) platforms, such as HiSeq 2000. This process provides a large number

of SNPs to develop a high-density genetic map. Moreover, methods such as restriction site-associated DNA sequencing (RAD-seq) (Baird *et al.*, 2008) allow labeling of fragments with barcode sequences and pooling of several dozens of samples to one library, thus markedly reducing the per-sample cost and producing results in a reasonable time. Therefore, these NGS-based methods have been widely used to develop high-density genetic maps in plants (Xie *et al.*, 2010; Chutimanitsakun *et al.*, 2011; Pfender *et al.*, 2011; Jia *et al.*, 2013; Saintenac *et al.*, 2013; Zhang *et al.*, 2013; Chen *et al.*, 2014).

Grain size or weight is genetically controlled by multiple genes, and a large number of QTLs for grain traits have been characterized in wheat in the past two decades (Gegas *et al.*, 2010; Tsilo *et al.*, 2010; Prashant *et al.*, 2012; Maphosa *et al.*, 2014; Rasheed *et al.*, 2014; Williams and Sorrells, 2014; Kumar *et al.*, 2016; Brinton *et al.*, 2017). The identified QTLs are distributed along all the wheat chromosomes, especially the stable and major QTLs on A subgenome 1A (Gegas *et al.*, 2010; Williams and Sorrells, 2014), 2A (Tsilo *et al.*, 2010), 3A (Gegas *et al.*, 2010; Kumar *et al.*, 2016), 4A (Prashant *et al.*, 2012), 5A (Gegas *et al.*, 2010; Williams and Sorrells, 2014; Brinton *et al.*, 2017), 6A (Gegas *et al.*, 2010; Maphosa *et al.*, 2014), and 7A (Tsilo *et al.*, 2010; Kumar *et al.*, 2016). Of those characterized QTLs, one QTL for grain weight on 5A was further validated with two near-isogenic lines and fine-mapped to a genetic interval of 4.32 cM corresponding to 74.6 Mb of genomic sequence (Brinton *et al.*, 2017). However, it remains a challenge to identify the causative genes in such a long genomic interval due to functional redundancy (genetic buffering) of genes in three homoeologous genomes and the highly repetitive nature of the wheat genome (Slade *et al.*, 2005; International Wheat Genome Sequencing Consortium, 2014). Until recently, most candidate genes for grain size and weight in wheat were characterized by homology-based cloning (Maphosa *et al.*, 2014; Kumar *et al.*, 2016). Several genes in rice have been proven to influence wheat grain size and weight, such as *GW2* (*TaGW2*), *GS3* (*TaGS-D1*), *CKX2* (*TaCKX6*), *GS5* (*TaGS5*), *TGW6* (*TaTGW6*), *GASR7* (*TaGASR7*), and *GIF1* (*TaCWI*) (Li and Yang, 2017). Moreover, the starch and sucrose biosynthesis pathway, including the genes *TaAGPL*, *TaAGPS*, and *TaSus2*, which was unraveled in model species, was suggested to function in wheat (Jiang *et al.*, 2011; Hou *et al.*, 2017). Thus, reverse genetics is an efficient approach to characterize the underlying genetic components of morphogenesis in wheat (Li and Yang, 2017). Nevertheless, along with rapid progress in genome sequencing, forward genetics would greatly facilitate the characterization of candidate genes involved in the development of wheat grain size and weight.

Einkorn wheat, *Triticum monococcum* ssp. *monococcum* L. ($A^m A^m$, $2n=2x=14$), the only cultivated diploid wheat, was domesticated from its wild species, *T. monococcum* ssp. *boeoticum* ($A^b A^b$). As one of the remaining unaltered crops, wild einkorn wheat grew in a natural environment without selection for thousands of years (Jing *et al.*, 2007). This species preserves a large number of phenotypic variations that would facilitate the dissection of genetic architectures for important agronomic traits (Jing *et al.*,

2007; International Wheat Genome Sequencing Consortium, 2014; Zaharieva and Monneveux, 2014). The leaf rust-resistant gene *Lr10*, the most important domestication gene *Q*, and the vernalization genes *Vrn1* and *Vrn2* were cloned with the help of einkorn wheat (Stein *et al.*, 2000; Feuillet *et al.*, 2003; Yan *et al.*, 2003, 2004; Simons *et al.*, 2006). Thus, the genome characteristics, high polymorphism, and diversified traits make einkorn wheat a good model for gene discovery and breeding improvement in hexaploid wheat (*Triticum aestivum*, $2n=6x=42$, AABBDD) (Stein *et al.*, 2000; Shindo *et al.*, 2002; Yan *et al.*, 2003).

To unravel the genetic architecture of grain traits in einkorn wheat, a recombinant inbred line (RIL) population of wild and cultivated einkorn wheat was explored to map QTLs and characterize the underlying candidate genes. We exploited population SNPs through a RAD-seq approach, developed a high-density genetic map, conducted a genome-wide QTL analysis for grain traits, and elucidated the candidate genes or gene pathways underlying the QTLs based on comparative genomics and RNA sequencing (RNA-seq) analysis. The data revealed complex genetic components determining grain size variation and positive alleles retained across domestication in einkorn wheat. The whole-genome transcriptomic profiling further elucidated the candidate genes underlying QTLs with significantly differential expression between cultivated and wild einkorn wheat. Furthermore, the superior alleles identified in this work provide opportunities for genetic improvement of wheat.

Materials and methods

Plant material and phenotyping

The 109 RIL population (F_{10}) of *T. monococcum* ssp. *boeoticum* (KT1-1) \times *T. monococcum* ssp. *monococcum* (KT3-5) was selected for linkage map construction and QTL mapping. The population was kindly provided by the KOMUGI Wheat Genetic Resources Databases of Japan and further developed in our laboratory. The growth habit of KT1-1 is winter type and that of KT3-5 is spring type. Field trials with two replicates were conducted in a randomized complete block design at Beijing in 2010–2011 (E1, Beijing2011), 2011–2012 (E2, Beijing2012), 2012–2013 (E3, Beijing2013), and 2013–2014 (E4, Beijing2015), and at Zhengzhou in 2013–2014 (E5, Zhengzhou2014) growing seasons, as previously reported (Yu *et al.*, 2017).

Each RIL, along with the parents, was planted in 2 m long rows spaced 40 cm apart, with 20 cm between individual plants within each row. The seeds were harvested from five randomly selected representative plants from each line and replicate, and then manually threshed. Grain weight was determined using 100 grains with three replicates, and transformed to thousand-grain weight (TGW, g). From each replicate, 100 grains for each RIL were imaged and processed using SC-G software (WSeen, Hangzhou, China), and the average values of grain length (GL, mm), grain width (GW, mm), grain length/width (GLW), grain area (GA, mm^2), and grain circumference (GC, mm) were calculated. TGW was determined for grains from all five environments, while the grain size traits were collected from E2 to E5.

RAD library construction and sequencing

Genomic DNA was isolated from young leaves of 14-day-old seedlings using the cetyltrimethylammonium bromide protocol (Saghai-Marouf *et al.*, 1984). Complexity-reduced genomic libraries prepared using the restriction endonuclease *SbfI* (CCTGCAGG) have been reported in

other species with large genomes (Chutimanitsakun *et al.*, 2011; Pfender *et al.*, 2011). The genomic DNA of RILs was sufficiently digested with *SbfI* and processed into RAD libraries according to the protocol of Baird *et al.* (2008). We used a set of 16 barcoded adapters with sticky ends complementary to the 3' overhang (TGCA-3') created by *SbfI*. RAD libraries with an average size of 500 bp were constructed. For each library, 16 RILs were pooled together with each 6 bp barcode sequence to distinguish them and loaded on to one lane, except for the seventh lane, which contained 13 RILs and two parental lines. The RAD libraries were sequenced for a single read (101 bp) on an Illumina Genome Analyzer IIx (Illumina, San Diego, USA).

RAD-seq data processing and SNP calling

RAD-seq data were processed by commands from Stacks v0.98 (Catchen *et al.*, 2011). First, raw data were split into individuals according to the first 6 bp barcode sequences of reads and filtered using sliding window methods by *process_radtags* with parameters of '-e sbfI -c -q -r -i fastq -E phred33'. Any reads containing uncalled bases or phred33 quality scores less than 10 in any sliding window of $0.15\times$ read length were removed and discarded. Then, SNP calling from these tag sequences with the *SbfI* site was performed by *ustacks*, *cstacks*, and *sstacks*. Finally, SNPs were transformed to genotypes, filtered with a calling ratio $>40\%$, and applied to construct a linkage map.

Genetic map construction

All collected genotypes for the RILs were subjected to linkage mapping, and the distorted markers (Chi-square test $P < 0.01$, deviating from the expected 1:1 Mendelian segregation ratio) were excluded if these markers greatly affected the order of their neighbor markers or excessively changed genetic distance. Linkage grouping and marker ordering were conducted using Joinmap V4.0 (Van Ooijen and Voorrips, 2004) based on logarithm of the odds (LOD) scores ranging from 3.0 to 15.0 and MSTmap (Wu *et al.*, 2008), respectively. Recombination frequencies were converted into cM using the Kosambi function (Kosambi, 1943). The final graphical linkage maps were generated using MapChart2.0 (Voorrips, 2002).

Syntenic analysis

BLASTN was used to align the SNP markers against the physical maps of hexaploid wheat (IWGSC WGA v0.4; accessed from <https://urgi.versailles.inra.fr/>; last accessed June 9, 2019) and barley (IBSC RefSeqv1.0) (Mascher *et al.*, 2017). We filtered the BLASTN output data based on e -value $< 1e^{-10}$ and query coverage $\geq 90\%$. Hits that did not meet the conditions were discarded. To balance the relationship bias of alignment in different genomes, different percentage identity thresholds were set: 91% for the H genome of barley, and 98%, 96%, and 96% for the A, B, and D genomes of wheat, respectively. The cleaned alignment data were used to compare the genetic map with the physical maps using OmicCircos (Hu *et al.*, 2014) implemented in R (R Core Team, 2016).

Statistical analysis and QTL mapping

All statistical analysis was conducted in R (R Core Team, 2016). The broad-sense heritability (H^2) was estimated by using analysis of variance (ANOVA), and Pearson's correlation coefficients among traits were calculated. The coefficient of variation was independently calculated for each individual environment as σ/μ , where σ represents the standard deviation and μ represents the mean of the phenotypic data in the population. QTL analysis was performed using the composite interval mapping (CIM) method in Windows QTL Cartographer V2.5 (Wang *et al.*, 2012b) as previously described (Yu *et al.*, 2017). QTLs were reported according to previous methods (Kumar *et al.*, 2016; Yu *et al.*, 2017), and QTLs detected from at least two environments or linked with other trait QTLs or homologously mapped genes were also retained. QTLs linked with flanking markers or overlapping confidence intervals (CIs; ± 1 LOD) were considered as one QTL for each trait with the CI reassigned by

the overlapping genetic positions, while unique genomic regions were considered to be regions with at least one QTL included. The total explained variance by QTLs was estimated according to our previous study (Yu *et al.*, 2017).

RNA-seq and data analysis

KT1-1 and KT3-5 seedlings at the one-leaf stage were grown in pots (8 cm²×8 cm) with nine seedlings per pot under 4–6 °C (8 h light) for 6 weeks. The seedlings were then transplanted to 12 pots (25 cm²×25 cm) with four plants in each pot. They were grown in a greenhouse under conditions of 16 h light at 25 °C and 8 h darkness at 15 °C. The flowering time was recorded for each plant. After removal of their top and base two or three spikelets, the spikes were harvested at 0, 7, 14, and 21 days after flowering (DAF), starting from the anther emerging in the middle of the spike; three biological replicates were sampled, and three primary spikes from different plants in the same pot were collected for each replicate. For each line, nine primary spikes were collected from different plants at the end of the daylight period for each developmental stage and immediately frozen in liquid nitrogen and stored at –80 °C. Three primary spikes for each replicate were pooled for RNA isolation. Total RNA extraction and quality assessment were conducted as previously reported (Zhang *et al.*, 2015). Library construction and RNA-seq were performed by BGI (Shenzhen, China). cDNA libraries with an average insert size of 300 bp from 24 samples were prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina San Diego, USA) and sequenced on a HiSeq4000 (Illumina, San Diego, USA) according to the manufacturer's standard protocols.

The raw reads were quality filtered for adaptor contamination and low-quality or unknown nucleotides. The resultant clean reads were aligned against the wheat accession Chinese Spring TGACv1.32 genome assembly (ftp://ftp.ensemblgenomes.org/pub/plants/release-32/fasta/triticum_aestivum/; last accessed June 9, 2019). Transcript count information for each gene sequence was calculated and normalized to the fragments per kilobase of transcript per million mapped reads (FPKM) values (Trapnell *et al.*, 2010). Significantly differentially expressed genes (DEGs) were screened using the bioconductor package NOISeq (Tarazona *et al.*, 2011). Genes with $|\log_2(\text{FPKM}_{\text{KT3-5}}/\text{FPKM}_{\text{KT1-1}})| > 1$ and probability > 0.8 were identified as DEGs, while genes with probability > 0.7 were considered as suggestive DEGs. Gene functions were assigned according to the best match of the alignments to the Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/>; last accessed June 9, 2019), NR (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>; last accessed June 9, 2019), and Gene Ontology (GO) databases using BLASTP (v2.2.23) with default parameters. GO terms and KEGG pathways of the investigated genes were extracted. GO and KEGG pathway enrichment analysis was performed using the phyper function implemented in R (R Core Team, 2016). GO terms with corrected *P* values ≤ 0.05 and KEGG pathways with *Q* values ≤ 0.05 were defined as significant enrichments.

Results

Multigenic control of grain size traits in einkorn wheat

The grain size traits and TGWs for the 109 RILs and their parents were determined (Fig. 1; Supplementary Table S1 and Fig. S1 at JXB online). The wild einkorn wheat KT1-1 generally had a small seed size (GL=6.40 mm and GW=1.59 mm), while the cultivated einkorn KT3-5 had larger seeds (GL=7.59 mm and GW=2.60 mm in E2). Among the RIL population, GL and GW were continuously distributed and preserved a transgressive inheritance; for example, GL varied from 6.17 to 8.85 mm with a mean of 7.71 mm in E2. High heritability (H^2) was observed in both traits (0.86 for GL and 0.82 for GW). However, significantly higher coefficients of variation were documented

in GW across four environments (10.37% of GW versus 7.36% of GL; $P=0.0094$ based on *t*-test), demonstrating that GW had larger variations. Moreover, TGW showed the highest coefficient of variation across all environments, from 21.16% (E1) to 32.64% (E3), but with high heritability (0.85); for other traits, values of H^2 were not lower than 0.8 (Supplementary Table S1). Thus, the large variation of the observed traits was proposed to be mainly under genetic control with multiple loci in this RIL population.

Significant correlations were observed among GL, GW, GLW, GA, GC, and TGW (Supplementary Table S2). GL had the highest positive correlation with GC ($r=0.99$, Bonferroni-adjusted $P<0.01$), followed by TGW versus GA, GC versus GA, and GW versus GA (Fig. 1). TGW was positively correlated with all the grain-size-related traits except GLW (–0.33), and had a stronger correlation with GW (range 0.73 to 0.93, mean 0.88) than GL (range 0.66 to 0.82, mean 0.78) across all surveyed environments (Fig. 1; Supplementary Table S2). GLW was significantly negatively correlated with GW (–0.67), but showed no significant correlation with GL (0.24, $P=0.19$). GC was correlated more strongly with GL (0.99) than GW (0.63), while GA had an almost equal significant positive correlation with GL and GW (0.86 and 0.89, respectively). Thus, GW was a more important determinant of TGW, and GL and GW contributed differentially to the composite traits of GLW (grain shape), GA and GC (grain size) in this RIL population. Principal component analysis also identified two extracted principal components, PC1 for grain size (65.5–77.8%) and PC2 for grain shape (20.8–31.5%), in four environments (Supplementary Table S3; Supplementary Fig. S2).

Development of a high-density genetic map through the RAD-seq approach

Seven *Sbf*I reduced-representation libraries were constructed, and a total of 64.88 Gb of data, with 642 million reads, was subsequently generated. All reads were demultiplexed to the samples and subjected to quality control, and this resulted in $\sim 3.95 \pm 1.06$ million (mean \pm SD) reads per sample for further analysis (Supplementary Table S4). Genomic variation calling resulted in 42 278 putative SNPs at 25 805 genomic tags (11.36% of the total 227 244 tags). Of these, 25 609 SNPs distributed at 16 566 tags were retained with $> 40\%$ population calling rate, and these tags are hereinafter referred to as SNP markers for SNPs in one tag forming one haplotype.

These 16 566 SNP markers were used to construct the genetic map, along with 939 genetic markers from a previous study (Yu *et al.*, 2017). The resulting genetic map contained 10 876 genetic markers distributed on seven chromosomes, designated as Tm1A to Tm7A, based on the known mapped markers, and these markers were grouped into 1551 unique bins (Table 1; Fig. 2). The genetic map spanned 1873.04 cM with an average marker interval of 0.17 cM and average marker density of 5.8 per cM (Table 1). The number of markers on each chromosome varied from 1343 (Tm1A) to 1732 (Tm2A), and the number of genetic bins varied from 176 (Tm4A) to 265 (Tm7A) per chromosome. The average bin length varied from 1.07 cM (Tm7A) to 1.50 cM (Tm4A).

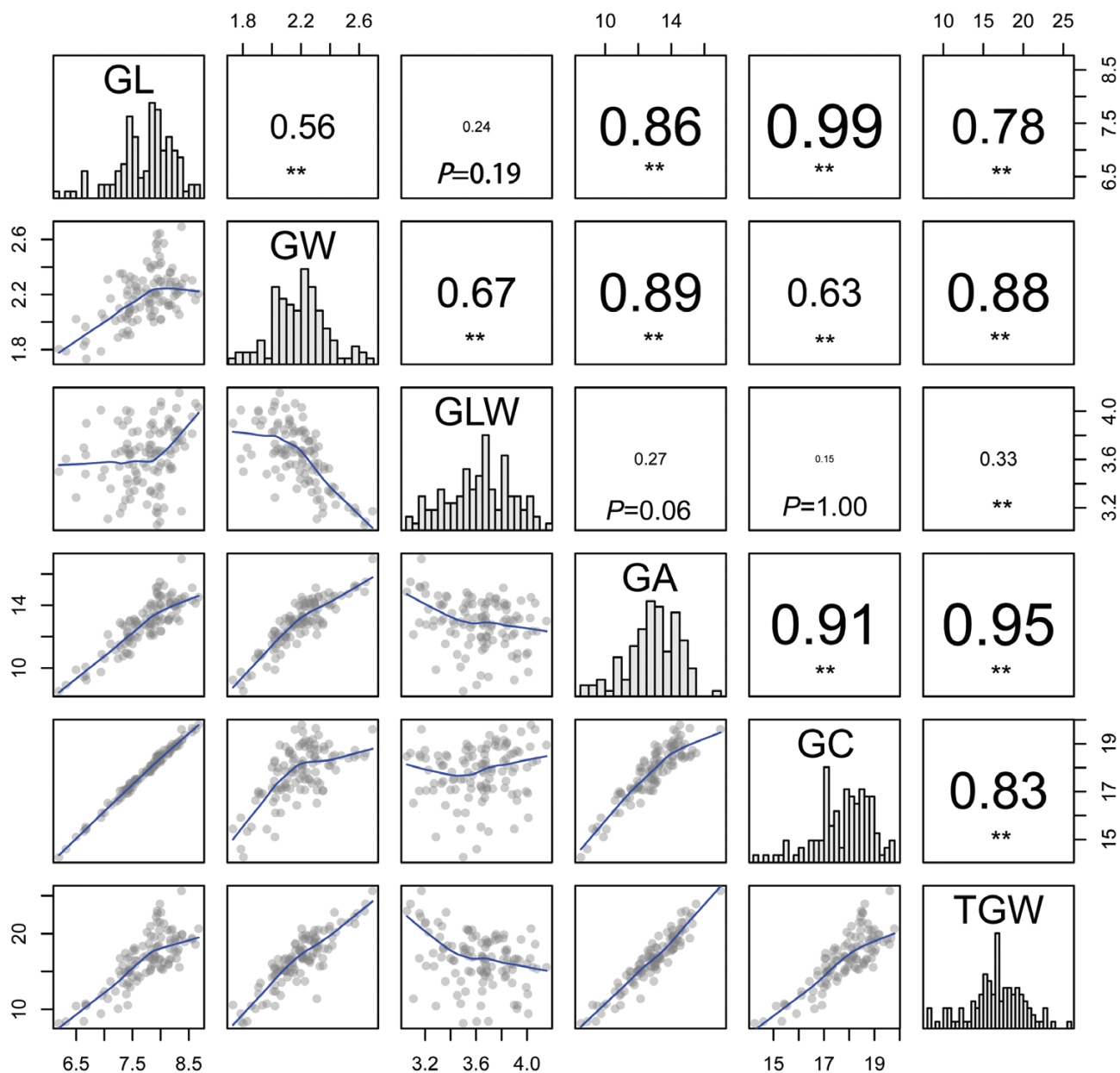


Fig. 1. Phenotypic performances, distribution, and correlation coefficients for six quantitative traits of parents and RILs, using the average phenotypic data. The frequency distribution of the phenotypic data for each trait is shown in the histograms. The X-Y scatter plots in the lower left panels show the correlations between traits, while the corresponding Pearson's correlation coefficients and *P* values of multiple comparison tests are located in the upper right panels. **P*<0.05, ***P*<0.01. GL, grain length (mm); GW, grain width (mm); GLW, grain length/width; GA, grain area (mm²); GC, grain circumference (mm); TGW, thousand-grain weight (g).

Table 1. Summary information of the high-density einkorn wheat genetic map

Chromosome	Length (cM)	SNPs	Other markers	Total number of markers	Bin number	Bin length (cM)	Marker interval (cM)	Maximum gap (cM)
Tm1A	245.33	1233	110	1343	206	1.19	0.18	4.85
Tm2A	265.16	1621	111	1732	230	1.15	0.15	6.71
Tm3A	293.50	1466	190	1656	242	1.21	0.18	6.28
Tm4A	264.33	1453	103	1556	176	1.50	0.17	20.82
Tm5A	283.08	1269	153	1422	245	1.16	0.20	8.16
Tm6A	238.61	1357	115	1472	187	1.28	0.16	5.02
Tm7A	283.04	1538	157	1695	265	1.07	0.17	12.17
Total	1873.04	9937	939	10 876	1551	1.21	0.17	20.82

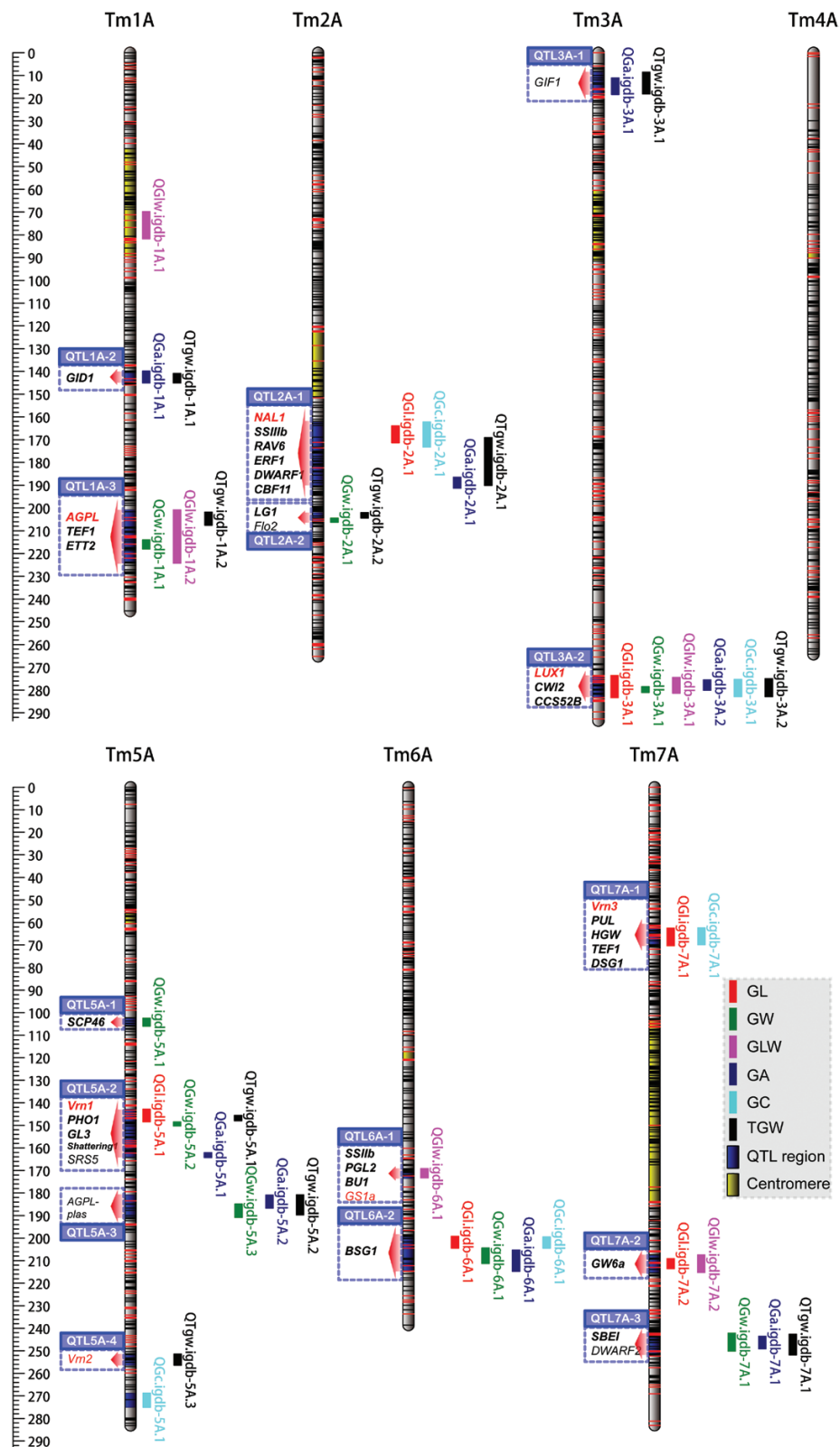


Fig. 2. QTLs detected genome-wide using the high-density genetic map of einkorn wheat. The genetic map showed 17 genomic regions harboring QTLs for six grain traits in an einkorn wheat RIL population of *T. monococcum* ssp. *boeoticum* (KT1-1) and *T. monococcum* ssp. *monococcum* (KT3-5). The detected QTLs for each trait from each environment were combined with confidence intervals and mapped on the genetic map. At each linkage group, each QTL is plotted on the right side, while the candidate genes in each QTL region are indicated on the left side. Detailed information on the QTLs is provided in Table 2. The candidate genes of each QTL region are shaded blue, and the red arrows show the QTL regions. Genes in red text were mapped through developing functional markers and genetic mapping, while black bold text denotes genes located inside the QTL region, and red or black normal text denotes genes surrounding the QTL region. The yellow shaded portions of each linkage group are the probable centromere regions. The positions of SNPs and other types of markers are denoted with black and red horizontal lines, respectively. GL, grain length; GW, grain width; GLW, grain length/width; GA, grain area; GC, grain circumference; TGW, thousand-grain weight.

Compared with a previous map constructed with the same population (Yu *et al.*, 2017), this high-density genetic map has been extended by approximately 496 cM (from 1377 cM to 1873 cM). The larger genetic map resulted primarily from additional intrachromosomal recombination (~439 cM) detected by the new mapped SNP markers. Meanwhile, 84 SNP markers located beyond other markers at the distal ends of chromosomes (of which 21 were mapped on Tm1AS/L, six on Tm2AS, two on Tm4AL, 51 on Tm5AS/L, and four on Tm6AS) covered 29 bins with ~57 cM. Regarding gaps in this high-density genetic map, all intervals were <10 cM between two neighboring markers, other than one gap on Tm4A (20.82 cM between bin3 and bin4) and two gaps on Tm7A (12.17 cM between bin263 and bin264, 10.02 cM between bin173 and bin174) (Fig. 2; Supplementary Table S5). The gaps detected here should be caused by lack of recombination in these intervals, not only in this RIL population (Shindo *et al.*, 2002; Hori *et al.*, 2007; Yu *et al.*, 2017) but also in other einkorn wheat populations (Jing *et al.*, 2007; Singh *et al.*, 2007). Thus, this genetic map was greatly improved in terms of marker density and evenness, and represents a high-quality map with thousands of markers and limited gaps.

Furthermore, the quality and genome coverage of the genetic map was evaluated with 9937 SNP markers against the barley (IBSC RefSeqv1.0) (Mascher *et al.*, 2017) and hexaploid wheat (IWGSC WGA v0.4) genomes. In total, 1834, 2187, 1693, and 922 hits were retained for the A, B, D, and H genomes, respectively, after alignment filtration. These alignments corresponded to 3102 SNP markers, and 2758 markers were mapped to the expected homologous chromosomes and showed high collinearity [average Spearman's rank correlation coefficient (ρ)=0.76] except the Tm4A-Ta4A and Tm7A-Ta7B comparisons (Fig. 3A). The 4AL/5AL/7BS translocations (Devos *et al.*, 1995) were observed with a genetic distance of ~50 cM in Tm5A and physical lengths of ~35 Mb, ~28 Mb, and ~20 Mb in Ta4B, Ta4D, and Hv4H, respectively, and two SNPs from Tm4A covering ~1 cM mapped to Ta7BS (Fig. 3B, C; Supplementary Fig. S3A, B). It is noteworthy that the mapped SNP markers cover >97% of all the chromosomes, except for Ta4A (92%) and Ta3B (93%) (Supplementary Fig. S3C). Therefore, the high collinearity provides syntenic blocks between einkorn wheat and available genomes, which would facilitate the identification of interesting genomic regions for further analysis.

Genetic architectures of grain size traits in einkorn wheat

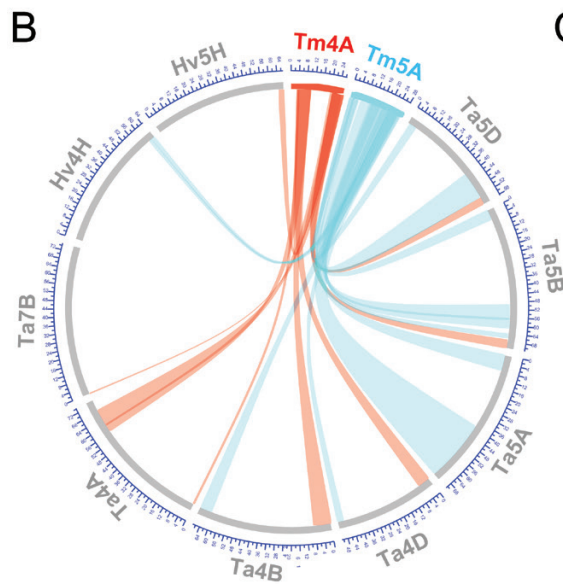
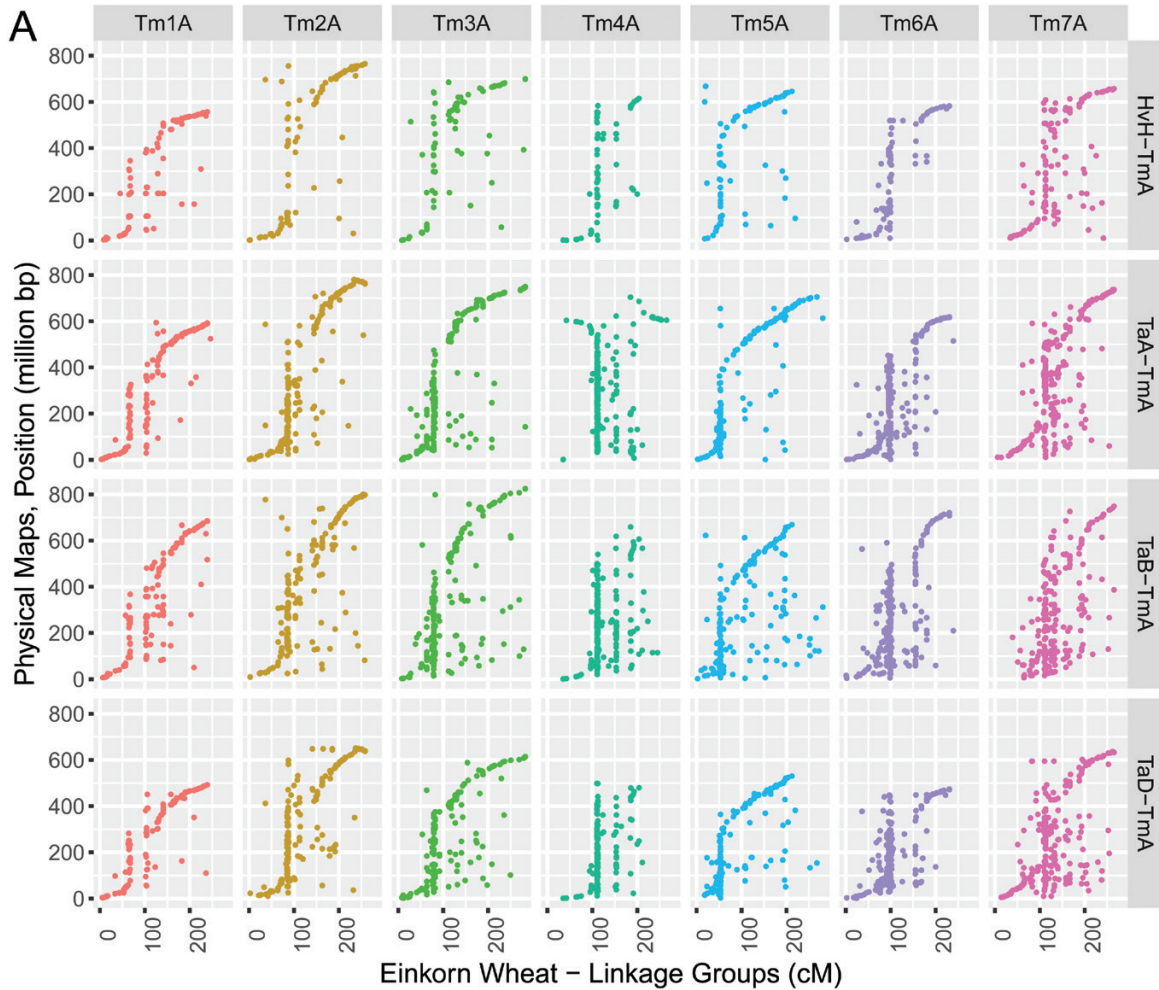
Using the CIM method, a total of 42 additive QTLs were identified and distributed across six chromosomes (except Tm4A); they had a LOD peak score of >3.4 and explained 6.4–38.1% of the phenotypic variation (Supplementary Fig. S4; Supplementary Table S6). Among the QTLs, 31 (74%) loci involved positive alleles from KT3–5 for increasing phenotypic values, while the other 11 (26%) had positive alleles from KT1–1 for increasing phenotypic values, suggesting that positive alleles for grain-size-related traits were present even in the parent with low phenotypic values. These

42 unique QTLs were assigned to 17 genomic regions for QTLs of several traits co-located at the same chromosomal region. This resulted in an average number of 2.5 QTLs per region, of which 3A-2 (273.6–282.9 cM) harbored the most QTLs, which were related to all six phenotypic traits (Table 2; Fig. 2).

To investigate the candidate genes underlying QTLs, homologous genes with functions related to grain size or weight reported in rice, barley, and wheat were retrieved and mapped to this genetic map. Except 1A-1, which was homologous to the chromosomal centromeric region, the remaining 16 QTL regions had homologous blocks in the hexaploid wheat and barley genomes. These syntenic blocks had average physical lengths of 24.29 Mb and 18.85 Mb in the hexaploid wheat A and barley H genomes, respectively (Supplementary Fig. S5). This process allowed 41 collected genes to be mapped, 40 in the wheat genome and 36 in the barley genome (Supplementary Table S7). Among these genes, 30 were mapped in the QTL confidence intervals, and seven (*Flo2*, *GIF1*, *SRS5*, *AGPL-plas*, *Vrn2*, *GS1a*, and *DWARF2*) were closely linked with their target QTL; however, four genes had a genetic distance >10 cM from the identified QTL, corresponding to five loci: *Sus2* (83.67 cM) and *GW7* (75.70 cM) on Tm2A, and *Sus1* (32.65 cM), *Sus2* (83.67 cM), and *GASR7* (103.17 cM) on Tm7A (Supplementary Tables S5 and S7). To confirm their concordant locations, *AGPL*, *Sus1*, *Sus2*, *Vrn1*, *Vrn2*, *Vrn3*, *NAL1*, *GS1a*, *GASR7*, and *GW7* were mapped by developing polymorphic markers (Supplementary Table S8), such as an insertion–deletion (InDel) marker for *Vrn3* (Supplementary Fig. S6). Collectively, *AGPL*, *Vrn1*, *Vrn3*, and *NAL1* were mapped to 1A-3, 5A-2, 7A-1, and 2A-1, respectively, and *Vrn2* and *GS1a* were located to within 5 cM of 5A-2 and 6A-1, which is consistent with the comparative homologous regions.

GL and GW QTLs

In total, 11 genomic regions contained QTLs for GL and GW (Supplementary Fig. S7). For GL, six QTLs were distributed over chromosomes Tm2A, Tm3A, Tm5A, Tm6A, and Tm7A, explaining 7.27–35.43% of the phenotypic variation (Supplementary Table S6). The KT1-1 alleles on Tm2A, Tm3A, Tm5A, and Tm6A decreased GL by 0.14–0.45 mm, while the allele on Tm7A increased GL by 0.16–0.34 mm. All detected QTLs represented by peak markers could explain 59.1% of the total phenotypic variation (Supplementary Table S9). Multiple comparison showed that when RILs inherited two or three positive alleles that would increase phenotypic values, GL increased significantly ($P<0.01$) (Supplementary Fig. S8; Supplementary Table S10). Eight GW QTLs were detected across all six QTL-located chromosomes, jointly explaining 56.8% of the total phenotypic variation (Supplementary Table S9). Most of these QTLs showed a negative effect of KT1-1 alleles, decreasing GW by 0.07–0.13 mm, while only *QGw.igdb-7A.1*, mapped in 7A-3 (242.2–250.2 cM), increased GW by 0.07 mm (Table 2). Three regions containing QTLs for both GL and GW, 3A-2, 5A-2, and 6A-2, were located on multiple genes, including *TmLUX1*, *CWI2*, and *CCS52B* for 3A-2, *Vrn1*, *PHO1*, *Shattering1*, and *GL3* on 5A-2, and *BSG1* on 6A-2 (Fig. 2). For GL, 2A-1 contained *NAL1*, 7A-1 contained



C

Genetic regions on einkorn wheat linkage map			Homologous blocks on hexaploid wheat and barley genomes			Evolution events
Chr.	Start (cM)	End (cM)	Chr.	Start (Mbp)	End (Mbp)	
Tm4A	33.88	38.17	4A	0.193	0.516	
	33.88	96.80	4D	0.464	56.725	
	38.17	96.80	4B	0.202	83.305	
	41.20	95.38	4A	603.374	570.263	
	202.97	252.81	4A	685.749	605.027	
	213.64	264.33	5B	675.87	710.624	4A/L5AL
	213.64	262.87	5D	534.871	562.625	4A/L5AL
	213.64	264.33	5H	648.521	668.689	4A/L5AL
	261.92	262.87	7B	0.141	1.381	4A/L7BS
Tm5A	1.39	42.08	5A	2.311	76.326	
	1.39	17.23	5B	1.741	47.059	
	17.23	35.96	5D	14.118	43.912	
	59.23	266.97	5A	393.76	704.834	
	106.95	202.81	5D	393.179	523.377	
	117.73	153.61	5B	496.579	566.973	
	175.30	184.83	5B	566.185	609.935	
	196.03	202.81	5B	644.576	658.916	
	211.36	217.41	4H	624.63	618.677	4A/L5AL
	212.78	262.76	4B	615.773	668.365	4A/L5AL
	212.78	250.35	4D	484.269	509.434	4A/L5AL
	228.34	253.13	4H	634.479	644.855	4A/L5AL

Fig. 3. Genomic collinearity and chromosomal structure variation revealed by the high-density genetic map of einkorn wheat. (A) SNP markers were aligned against four einkorn wheat-related genomes (A, B, and D from hexaploid wheat, and H from barley), and the positions of the hit markers were compared with physical locations from the four genomes. (B) Comparisons of the marker positions on homologous groups 4, 5, and 7 elucidate 4A/L5AL/7BS translocations using the einkorn wheat genetic map. The detailed information is given in (C).

Vrn3, *PUL*, *HGW*, *TEF1*, and *DSG1*, and 7A-2 contained *GW6a*. For GW, 1A-3 contained *AGPL*, *TEF1*, and *ETT*, 7A-3 contained *SBE1*, and the remaining QTLs overlapped

with QTLs for other traits. Of these genes, *Vrn1*, *Vrn3*, *AGPL1*, *TmLUX1*, and *NAL1* were genetically mapped. Specifically, a 9 bp deletion in the *AGPL1* promoter region (~1 kb upstream)

Table 2. QTLs detected with the CIM method using the high-density einkorn wheat genetic map

Trait	QTL	Environment	Chromosome	Location (cM)	LOD	PVE (%)	Direction	Additive	QTL region	LOD threshold
Grain length (GL)	QG <i>i</i> . <i>igdb</i> -2A.1	E2, E4 (GA, GC, TGW)	2A	163.9–171.4	9.3	21.4	-	0.25	2A-1	3.31
	QG <i>i</i> . <i>igdb</i> -3A.1	E2, E3, E4, E5 (GW, GLW, GA, GC, TGW)	3A	273.6–283.3	3.5–14.2	7.3–35.4	-	0.14–0.45	3A-2	
Grain width (GW)	QG <i>i</i> . <i>igdb</i> -5A.1	E3 (GW, GA, TGW)	5A	142.8–148.5	3.5	8.1	-	0.18	5A-2	
	QG <i>i</i> . <i>igdb</i> -6A.1	E4 (GW, GA, GC)	6A	199.2–204.5	4.5	8.9	-	0.21	6A-2	
	QG <i>i</i> . <i>igdb</i> -7A.1	E4, E5, E3 (GC)	7A	62.5–70.4	4.0–9.0	9.0–19.6	+	0.16–0.34	7A-1	
	QG <i>i</i> . <i>igdb</i> -7A.2	E2, E4, E3, E5 (GLW)	7A	209.1–213.6	6.9	15.5	+	0.19	7A-2	
	QG <i>w</i> . <i>igdb</i> -1A.1	E2 (GLW, TGW)	1A	214.0–218.1	6.4	12.9	-	0.08	1A-3	3.39
	QG <i>w</i> . <i>igdb</i> -2A.1	E2, E5 (TGW)	2A	204.4–206.3	6.4	12.7	-	0.08	2A-2	
	QG <i>w</i> . <i>igdb</i> -3A.1	E3, E4 (GL, GLW, GA, GC, TGW)	3A	278.5–281.1	8.8	24.1	-	0.13	3A-2	
Grain length/ width (GLW)	QG <i>w</i> . <i>igdb</i> -5A.1	E4, E3 (Specific)	5A	102.5–106.0	6.1	16.2	-	0.10	5A-1	
	QG <i>w</i> . <i>igdb</i> -5A.2	E2 (GL, GA, TGW)	5A	148.5–150.3	3.9	7.2	-	0.07	5A-2	
	QG <i>w</i> . <i>igdb</i> -5A.3	E2 (GA, TGW)	5A	184.8–190.9	4.7	8.9	-	0.07	5A-3	
	QG <i>w</i> . <i>igdb</i> -6A.1	E4, E2 (GL, GA, GC)	6A	204.4–211.3	3.4	9.0	-	0.07	6A-2	
	QG <i>w</i> . <i>igdb</i> -7A.1	E4 (GA, TGW)	7A	242.2–250.2	3.5	8.8	+	0.07	7A-3	
	QG <i>w</i> . <i>igdb</i> -1A.1	E3, E4 (Specific)	1A	69.8–81.8	4.1–4.2	9.4–9.8	-	0.089–0.092	1A-1	3.36
	QG <i>w</i> . <i>igdb</i> -1A.2	E2, E4, E3 (GW, TGW)	1A	200.8–224.3	5.0–5.3	9.6–12.2	+	0.09–0.10	1A-3	
	QG <i>w</i> . <i>igdb</i> -3A.1	E5 (GL, GW, GA, GC, TGW)	3A	274.4–281.4	4.1	8.0	-	0.10	3A-2	
	QG <i>w</i> . <i>igdb</i> -6A.1	E2, E5 (Specific)	6A	169.2–173.3	9.6–13.5	19.0–34.0	+	0.13–0.21	6A-1	
	QG <i>w</i> . <i>igdb</i> -7A.1	E2, E3, E5 (GL)	7A	207.5–215.4	5.6–7.3	10.9–15.6	+	0.10–0.13	7A-2	
Grain area (GA)	QG <i>a</i> . <i>igdb</i> -1A.1	E4 (TGW)	1A	139.9–145.1	6.9	17.5	+	0.92	1A-2	3.36
	QG <i>a</i> . <i>igdb</i> -2A.1	E2, E3 (GL, GC, TGW)	2A	186.5–191.3	3.4–7.8	7.7–17.2	-	0.60–0.73	2A-1	
	QG <i>a</i> . <i>igdb</i> -3A.1	E4 (TGW)	3A	11.0–18.3	4.8	11.2	-	0.73	3A-1	
	QG <i>a</i> . <i>igdb</i> -3A.2	E2, E3, E4, E5 (GL, GW, GLW, GC, TGW)	3A	275.5–280.1	3.4–10.4	6.9–27.6	-	0.46–1.16	3A-2	
Grain circumference (GC)	QG <i>a</i> . <i>igdb</i> -5A.1	E2, E3 (GL, GW, TGW)	5A	162.1–164.3	9.6	22.3	-	0.81	5A-2	
	QG <i>a</i> . <i>igdb</i> -5A.2	E4, E5 (GW, TGW)	5A	181.0–186.7	4.1	9.3	-	0.65	5A-3	
	QG <i>a</i> . <i>igdb</i> -6A.1	E2 (GL, GW, GC)	6A	205.3–214.7	4.1	8.8	-	0.50	6A-2	
	QG <i>a</i> . <i>igdb</i> -7A.1	E4 (GW, TGW)	7A	243.5–249.1	3.4	6.4	+	0.55	7A-3	
	QG <i>c</i> . <i>igdb</i> -2A.1	E2, E4 (GL, GA, TGW)	2A	162.2–173.2	6.7	16.4	-	0.48	2A-1	3.30
	QG <i>c</i> . <i>igdb</i> -3A.1	E2, E3, E4, E5 (GL, GW, GLW, GA, TGW)	3A	275.3–282.9	5.6–13.8	12.5–34.4	-	0.40–0.98	3A-2	
Thousand-grain weight (TGW)	QG <i>c</i> . <i>igdb</i> -5A.1	E2, E3 (Specific)	5A	268.8–275.1	6.9	20.8	-	0.51	5A-5	
	QG <i>c</i> . <i>igdb</i> -6A.1	E4 (GL, GW, GA)	6A	199.4–204.5	4.8	9.6	-	0.48	6A-2	
	QG <i>c</i> . <i>igdb</i> -7A.1	E2, E3, E4, E5 (GL)	7A	62.3–69.9	4.0–8.9	8.9–19.6	+	0.37–0.75	7A-1	
	QT <i>gw</i> . <i>igdb</i> -1A.1	E4, E2 (GA)	1A	140.9–145.1	7.0	18.3	+	2.18	1A-2	3.32
	QT <i>gw</i> . <i>igdb</i> -1A.2	E1, E4 (GW, GLW)	1A	201.8–207.6	4.7	9.0	-	1.20	1A-3	
QT <i>gw</i> . <i>igdb</i> -2A.1	E2, E3, E5 (GL, GA, GC)	2A	169.1–190.1	5.7–7.3	11.5–16.9	-	1.47–2.00	2A-1		
QT <i>gw</i> . <i>igdb</i> -2A.2	E1 (GW)	2A	202.1–204.4	7.7	15.9	-	1.59	2A-2		

Table 2. Continued

Trait	QTL	Environment	Chromosome	Location (cM)	LOD	PVE (%)	Direction	Additive	QTL region	LOD threshold
	<i>QTgw.igdb-3A.1</i>	E1, E4, E2 (GA)	3A	8.5–18.1	3.6–4.2	6.7–10.3	-	1.04–1.65	3A-1	
	<i>QTgw.igdb-3A.2</i>	E2, E3, E4, E7, E5 (GL, GW, GLW, GA, GC)	3A	275.0–282.9	8.0–14.2	15.5–38.1	-	1.74–3.03	3A-2	
	<i>QTgw.igdb-5A.1</i>	E2 (GL, GW, GA)	5A	145.6–148.0	8.2	16.8	-	1.82	5A-2	
	<i>QTgw.igdb-5A.2</i>	E1, E4 (GW, GA)	5A	180.8–189.7	3.5–12.7	8.0–28.8	-	1.38–2.15	5A-3	
	<i>QTgw.igdb-5A.3</i>	E2 (Specific)	5A	251.4–256.4	5.7	10.8	-	1.41	5A-4	
	<i>QTgw.igdb-7A.1</i>	E4 (GW, GA)	7A	242.7–251.8	3.5	6.7	+	1.29	7A-3	

The phenotypic data were collected from five environments, E1, E2, E3, E4, and E5, which represent Beijing 2011, Beijing 2012, Beijing 2013, Beijing 2014, and Zhengzhou 2014, respectively. Environments in italic text denote that QTLs in these environments were detected above 2.0 LOD but below threshold LOD scores (~3.3). The QTLs detected in a single environment were also reported because of their overlap with QTLs for other traits or linking with homologous genes at these regions. The overlapped QTL-related traits and some genomic region-specific QTLs are given in brackets. The mapped locations were defined as 95% CI from identified environments. Chromosomes Tm1A to Tm7A are abbreviated to 1A to 7A. PVE, proportion variations explained by QTL. Direction, additive effect estimated of KT1-1 allele: + positive effect, - negative effect (positive effect of KT3-5 allele). Additive, the KT1-1 allele effect. QTL regions are defined by the ranges of all confidence intervals from all traits. LOD thresholds were calculated from 1000-time permutations for each trait-environment combination. The thresholds reported here are the average of all environments for each trait.

was observed in KT1-1, and a SNP changed an amino acid from S (KT3-5) to G (KT1-1), but other SNPs in the other five exons were synonymous mutations (Supplementary Fig. S9). Moreover, the increased positive alleles of the detected nine QTLs for GW showed more significant divergence between different groups than the alleles for GL (Supplementary Fig. S8; Supplementary Table S10).

GLW, GA, and GC QTLs

The composite traits GLW, GA, and GC were calculated from GL and GW, as they directly reflect grain size (GC and GA) and shape (GLW). The most significant QTL for GLW, *QGlu.igdb-6A.1*, could explain 19.0–34.0% of the phenotypic variation (Table 2). This QTL was specially mapped to 6A-1, which contains the genes *SSIIb*, *PGL2*, and *BU1*, which participates in starch biosynthesis and brassinosteroid signaling, associated with starch accumulation and grain weight and length. The GLW-specific QTL region was mapped to 1A-1 and is generally syntenic with a large proportion of the centromeric region, and little information regarding homologous genes was available. For GA, *QGa.igdb-1A.1* was located on 1A-2 (139.9–145.1 cM) and explained 17.54% of the phenotypic variation, with a positive effect with the KT1-1 allele, which contained no QTLs for GL and GW (Table 2). However, this QTL overlapped with *QTgw.igdb-1A.1* and a homologous gene of *GID1*, which interacts with DELLA protein in plants to control plant height (Jiang and Fu, 2007). Moreover, *QGa.igdb-3A.1* overlapped with *QTgw.igdb-3A.1* on 3A-1, which was linked with *GIF1* (Fig. 2). Four GC QTLs overlapped with GL QTLs, a finding that is consistent with the strong correlation between GC and GL ($r=0.99$). Overall, genetic overlaps were observed between the three composite traits and GL/GW, which were revealed by the genetic co-locations of GL/GW with GLW (three of five), GA (six of eight), and GC (four of five).

TGW QTLs

Ten QTLs for TGW were identified, on Tm1A, Tm2A, Tm3A, Tm5A, and Tm7A (Table 2). *QTgw.igdb-1A.1* and *QTgw.igdb-7A.1* showed positive effects with the KT1-1 allele, increasing TGW by 2.18 g and 1.29 g, respectively (Table 2). The remaining QTLs decreased TGW by 1.04–3.03 g and explained 6.71–38.05% of the phenotypic variation. Several previously mapped loci affected TGW or its related traits, such as *Vm1* (5A-2) and *Vm3* (7A-1); *Vm1* from KT1-1 delays heading but *Vm3* promotes heading (Yu et al., 2017). *QTgw.igdb-5A.3* was mapped between 251.4 cM and 256.4 cM; little information about this region is known except that *Vm2* (248.03 cM) is closely linked with it (Table 2). Interestingly, genes in the starch biosynthesis pathway, *AGPL*, *AGPL-plas*, *SSIIIb*, *PHO1*, and *SBEI*, which mapped to 1A-3, 5A-3, 2A-1, 5A-2, and 7A-3, respectively (Fig. 2), could decrease grain size by the KT1-1 allele, except 7A-3. With respect to five grain traits, nine of the TGW QTLs overlapped; of these, seven of the eight GA QTLs and six of the eight GW QTLs were co-located, while only three of the six GL QTLs and two GW-associated GLW QTLs were co-located (Supplementary Fig. S7). The data also demonstrated that grain size traits (especially GA and GW) were associated more positively with

TGW, while grain shape (GLW) had a negative effect on TGW for at least three QTLs, *QGlu.igdb-1A.2*, *QGlu.igdb-6A.1*, and *QGlu.igdb-7A.1*.

Environmental effects on QTLs

Environment-wide analysis showed that traits varied on phenotypic performance across the different environments, but they expressed similar patterns, except GLW (Supplementary Fig. S1). From all 42 QTLs, 26 (61.90%) were detected from at least two environments, out of which 80% of GC and GLW QTLs, 67% of GL QTLs, 60% of TGW QTLs, and 50% of GA and GW QTLs were detected. Meanwhile, 16 (38.10%) QTLs were detected in one environment but overlapped with QTLs for other traits (Table 2). Among the traits examined, GL showed the highest H^2 (0.86) and correlation between environments (average $r=0.65$), while GW showed the lowest ($H^2=0.82$ and $r=0.53$) (Supplementary Table S1). Moreover, ANOVA revealed that environment and genotype–environment interaction contributed to phenotypic variation, with variations explained by environment ranging from 3.39% (GLW) to 13.72% (TGW) (Supplementary Table S9). Therefore, our data reflect that the environmental factors had effects on phenotypic performance, thus further affecting QTL mapping.

Candidate genes underlying QTLs

To detect dynamic profiles of genes involved in grain development, RNA-seq was performed with spikes of two parents at four spike developmental stages, 0, 7, 14, and 21 DAF. After filtering out low-quality or adapter-contaminated reads, a total of 130.2 Gb clean data was harvested from 24 libraries of eight samples (two accessions \times four developmental stages), each with three biological replicates. After mapping against gene sets of the A genome of hexaploid wheat, 40.38% reads covering 87.77% (28 561/32 539) of total genes were shown to have unique positions, and these were subjected to further analysis.

DEGs between KT1-1 and KT3-5 were compared across the four developmental stages. A total of 4959 DEGs, including 2061 up-regulated and 2898 down-regulated genes, were identified with a threshold of probability >0.8 and $|\log_2(\text{FoldChange})| >1.0$ (Supplementary Fig. S10A). GO and KEGG pathway enrichment analysis revealed that these genes were involved in carbon fixation and metabolism, amino acid metabolism, and regulation of nucleotide metabolism-related enzymes (Supplementary Fig. S10 B, C; Supplementary Tables S11 and S12). The DEGs on the QTL regions were characterized through comparative transcriptional profiling along spike developmental processes. Of these mapped homologous genes, *SRS5* (Segami *et al.*, 2012) had FPKM values of 547 and 650 at 0 and 7 DAF in KT3-5, compared with 227 and 257 in KT1-1, and an overall up-regulation of 1.67- to 2.53-fold in KT3-5 was observed at all investigated stages (Fig. 4; Supplementary Table S13). Furthermore, *SRS5* was mapped on to 5A-2 (142.8–164.3 cM), which contributes to GL, GW, GA, and TGW (Fig. 2). Thus, *SRS5* may be a candidate gene for this QTL, affecting $\sim 6.22\%$ GA and $\sim 8.78\%$ TGW (Supplementary Table S6).

The gene encoding the ADP-glucose pyrophosphorylase (*AGPase*) large subunit, which is a rate-limiting enzyme that catalyzes the formation of ADP-glucose (ADPG), the substrate for starch biosynthesis (Georgelis *et al.*, 2007), was mapped to 1A-3 (200.8–224.3 cM) based on homology analysis. This gene (*AGPLcyto*, ID17), which is crucial for grain filling (Yang *et al.*, 2004), was highly expressed, with FPKM values of 398 and 246 at the middle and late stages, respectively (Supplementary Table S13). Furthermore, *AGPScyto* (ID18), encoding the *AGPase* small subunit, had a similar pattern with high expression levels (FPKM value reaching 1194 at 14 DAF). These two genes were differentially expressed between the two parental lines, at 7 DAF for *AGPLcyto* and 14 DAF for *AGPScyto* (probability >0.7) (Supplementary Table S13). The *AGPL* in einkorn wheat was sequenced, and several variations were detected along both the promoter and genic regions, including InDels and SNPs (Supplementary Fig. S9). A polymorphic marker based on an InDel on intron I was co-localized with 1A-3, further confirming the reliability of its homology-based mapping. Our data indicated that variation in the expression levels of its two subunits affected the formation of ADPG, and might further affect starch accumulation and even grain weight.

A total of 48 genes in the starch biosynthesis pathway were retrieved from the Chinese Spring database based on previous information (Krasileva *et al.*, 2017), and the expression patterns of their homologs in both parental lines were compared (Fig. 5A). In total, 31 genes (65%) were significantly differentially expressed in at least one stage with a probability >0.7 (Fig. 5B). The restricted enzyme gene, *ADPGT* (*ADPG Transporter*) or *BT1* (*Brittle1*), which transports ADPG from the cytoplasm to amyloplasts in cells (Sullivan *et al.*, 1991), was highly expressed in KT3-5 (FPKM >800) with fold changes of 1.67 and 4.86 (probability >0.7) for 7 DAF and 14 DAF, respectively. Another gene, *Starch Phosphorylases 1* (*PHO1*), which is responsible for the conversion of ADPG to the precursor for starch biosynthesis by starch synthase (SS, EC 2.4.1.1), was expressed in KT3-5 at a level more than 4-fold greater than in KT1-1 at 14 DAF. Moreover, *Sucrose synthase* (*Sus*, ID6) had much higher expression levels, with FPKM varying from 85 to 1505, especially at the late developmental stages; this could compensate for the expression of another five low-abundance *Sus* copies (ID1–5) (Fig. 5B; Supplementary Table S13). *UDP-glucose pyrophosphorylase* (*UGPase*, ID15), *Starch branching enzyme IIa* (*SBEIIa*, ID42), and *SBEIIb* (ID43) also showed similar patterns. Although several DEGs were expressed highly in KT1-1 at 0 or 7 DAF, for example, *Sus* (ID2–3), *Frk* (ID11), *UGPase* (ID16), *PHO2* (ID27–28), *GBSSI* (ID30), *SBEI* (ID39), *ISAIII* (ID47), and *PUL* (ID48), most of them were less strongly down-regulated or even up-regulated in KT3-5 at late developmental stages, which is a critical period for starch accumulation (Yang *et al.*, 2004). Thus, in the developing spikes, rate-limiting functional enzymes are very important for starch biosynthesis, and differential expression of relevant genes would affect starch accumulation and grain development.

Discussion

Grain size as a complex trait is poorly understood in hexaploid wheat. In this study, an einkorn wheat RIL population was utilized to reveal the genetic architecture of grain size

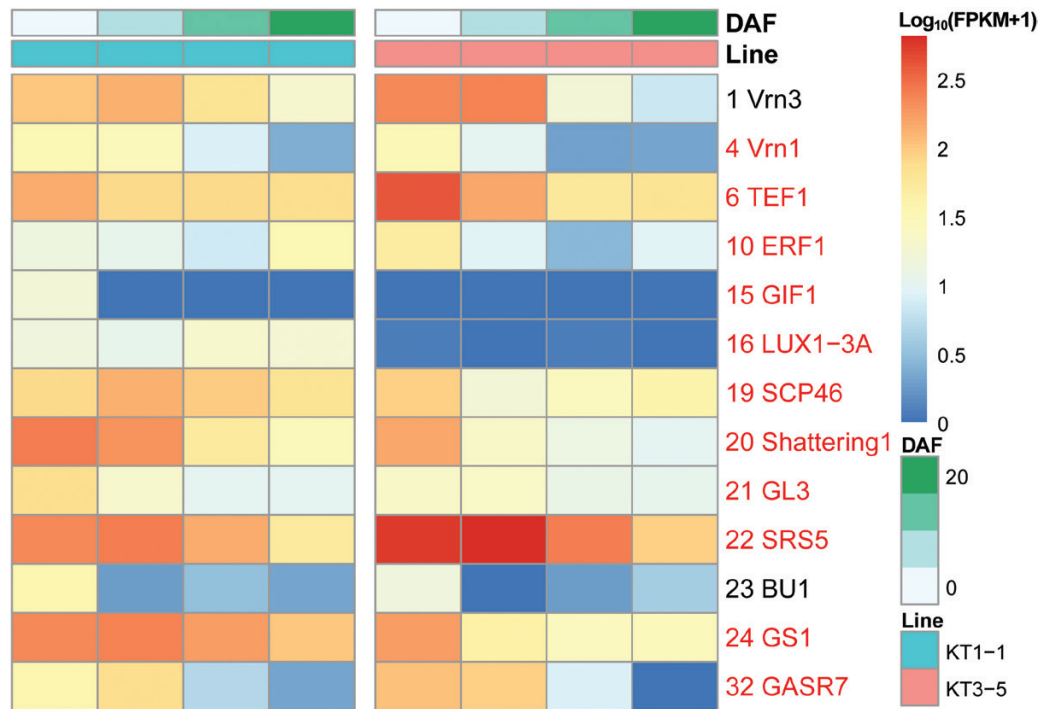


Fig. 4. Transcriptional profiles of genes mapped to QTL regions in two parental lines. Only genes differentially expressed in at least one developmental stage (probability >0.7) were retained; genes with names in red text had probability >0.8 from NOISeq. The $\log_{10}(\text{FPKM}+1)$ transformed data were plotted.

in a diploid progenitor. RAD-seq, combining NGS and restriction enzyme digestion to reduce genome complexity, was used to provide genetic polymorphisms at a genome scale to develop a high-density genetic map. This high-density map contains 10 876 evenly distributed genetic markers, had 1551 bins, and covered >97% of the wheat and barley genomes, demonstrating its good quality for einkorn wheat genetic and genomic research. In particular, comparative genomics could not only aid in examining the syntenic blocks with genomes of wheat relatives but also provide genomic sequences to dissect interesting regions and reveal structural variation between different genomes.

Novel QTLs were detected in einkorn wheat by comparison with tetraploid and hexaploid wheat populations

Grain size QTLs have been widely studied in wheat, and have been detected on all chromosomes of tetraploid and hexaploid wheat (Gegas et al., 2010; Tsilo et al., 2010; Peleg et al., 2011; Prashant et al., 2012; Maphosa et al., 2014; Rasheed et al., 2014; Russo et al., 2014; Williams and Sorrells, 2014; Golan et al., 2015; Wu et al., 2015; Kumar et al., 2016; Brinton et al., 2017; Cheng et al., 2017). However, very limited information is available for QTL analysis of grain size and weight in einkorn wheat. In this study, genome-wide QTL analysis was conducted and identified 17 genomic regions that contribute to grain size and weight. Based on syntenic regions between einkorn wheat and hexaploid wheat (Fig. 3) and marker information on the wheat genome (IWGSC RefSeq v1.0; https://urgi.versailles.inra.fr/jbrowseiwgsc/gmod_jbrowse/; last accessed June 9,

2019), QTLs that overlapped with markers from other studies were considered common regions; otherwise, QTLs were considered novel. In our study, five of 17 QTL regions were newly detected, including 1A-3, 5A-4, 5A-5, 6A-2, and 7A-3. The other 12 regions overlapped with QTLs or markers from other studies (Wang et al., 2009; Gegas et al., 2010; Peleg et al., 2011; Russo et al., 2014; Golan et al., 2015; Wu et al., 2015; Kumar et al., 2016; Brinton et al., 2017; Cheng et al., 2017). For example, a TGW QTL detected in the tetraploid wheat population linked with *wPt-7053* (Peleg et al., 2011), which is located on 676.40 Mb of 7A; similarly, QTL-27 (Kumar et al., 2016) and *QGl.cau-7A.3* (Wu et al., 2015) from hexaploid wheat were located at 674.27–705.13 Mb and 671.42–679.96 Mb, respectively, corresponding to 7A-2 (670.94–693.33 Mb) in this study. However, 7A-2 was detected to be associated with GL and GLW, while this locus only affected GL/GW/GA/TGW and GL in hexaploid wheat (Wu et al., 2015; Kumar et al., 2016). Another region, 1A-2, containing GA and TGW QTLs, was co-localized with the meta-QTL MQTL2, which linked with *Glu-1A* on Ta1A (Gegas et al., 2010). Three QTL regions, 3A-2, 5A-2, and 7A-1, involved 12 QTLs for grain size and weight. By comparing with our previous data, we were able to identify that these regions co-located with QTLs for heading date under the control of three genes, *TmLUX1*, *Vrn1*, and *Vrn3*, which are segregated in this RIL population (Yu et al., 2017). In addition, significant negative correlations between grain size traits (except GLW) and heading date were found, verifying the pleiotropism of these loci (Supplementary Table S2). Thus, this study suggests that the genes that control the initiation and duration of reproductive development would further affect the final performance of grain size and weight.

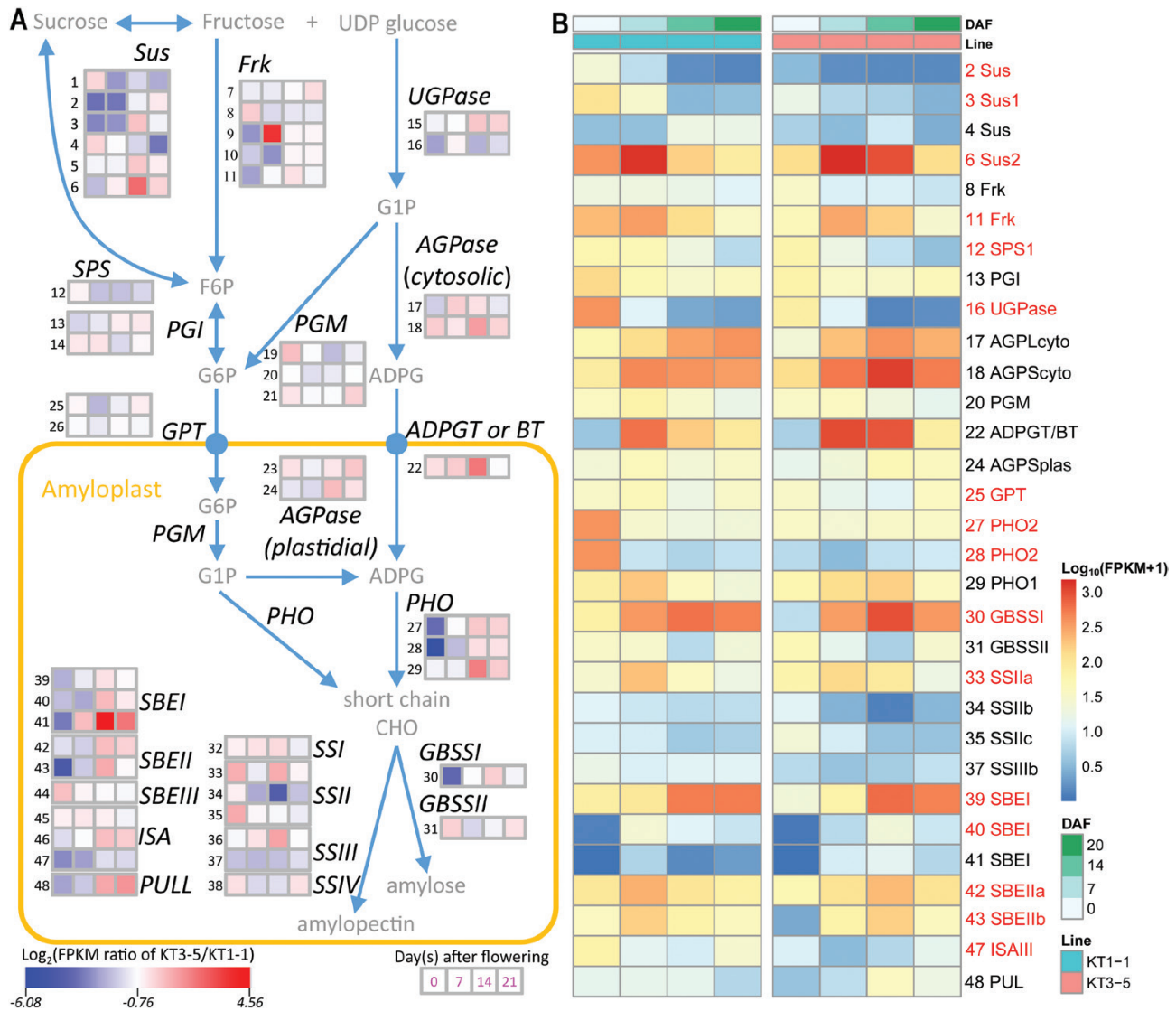


Fig. 5. Fold changes and expression patterns of starch biosynthesis genes across four spike developmental stages between two parental lines of einkorn wheat (KT1-1 and KT3-5). Only genes differentially expressed in at least one developmental stage (probability >0.7) were retained; genes with names in red text had probability >0.8 from NOISeq. (A) Wheat starch biosynthesis pathway. Heatmap of the log₂ fold changes of KT1-1 versus KT3-5 in FPKM at four developmental stages (0, 7, 14, and 21 DAF). (B) Heatmap of the expression profiles of starch biosynthesis pathway genes in grains of KT1-1 (left) and KT3-5 (right). The log₁₀(FPKM+1) transformed data are plotted.

Our data indicate that regulation of grain size and weight in einkorn wheat has a similar genetic basis to tetraploid and hexaploid wheat, but divergent functions of some loci that regulate grain size.

The environment had various effects on QTLs of different traits

Effects of environmental factors on QTL detection have been reported, with these factors causing QTLs not to be identified from every field trial or crop season (Kuchel *et al.*, 2007; Maphosa *et al.*, 2014; Tyagi *et al.*, 2015; Wu *et al.*, 2015). In our study, different traits showed different phenotypic plasticity, as revealed by the environmental contributions to phenotypic variation from ANOVA and QTL calling ratios from different environments. Using ANOVA, the environment was found to contribute 5.78% of the phenotypic variation of GW, which

was the largest among the traits except for TGW, indicating that GW is more liable to change as the environment changes in this population. Genotype–environment interactions were also explored (Supplementary Table S9), revealing that 3A-2 interacted with the environment for all six traits, 7A-1 for GL/GC, and 7A-3 for TGW/GW. These regions contained genes (*TmLUX1* and *Vrn3*) affecting heading date, which shows the environmental adaptation of einkorn wheat. The 3A-2 allele from KT3-5 could promote flowering and increase grain size and TGW, while the 7A-3 allele from KT1-1 showed a similar effect (Yu *et al.*, 2017). In addition, the environment interacted with the regions 2A-1 and 6A-2 for GA and GLW, respectively. Thus, environmental factors had a certain effect on the investigated traits of this RIL population, especially the development-related traits with large phenotypic plasticity. This might suggest that the influence of environmental factors on grain size traits should be considered and

evaluated for breeding improvement and the introduction of new varieties.

Candidate genes underlying QTLs were predicted based on comparative genomics and transcriptomics

The underlying candidate genes of genetic loci could help to understand morphogenesis and provide diverse alleles for breeding improvement in wheat. The past two decades have witnessed the characterization of several causative genes for grain size and weight in crops (Li and Yang, 2017). In this work, several genes have been shown to be the candidate genes for grain size in einkorn wheat by comparative analysis and transcriptomic profiling with RNA-seq. Those mapped homologous genes were differentially expressed at the early to middle stages of development for expanding grain volume before initiating grain filling in developing seeds. The homolog of rice *SRS5* showed successively higher expression in KT3-5 across all developmental stages. Mutant *SRS5* with an amino acid substitution reduced seed length by 1.38 mm by decreasing cell and lemma length, while the wild type could partially rescue the mutant phenotype in transgenic plants (Segami et al., 2012). Starch accumulation was determined to be critical to grain filling in wheat (Yang et al., 2004), and the starch biosynthesis pathway has been well characterized (map00500 in KEGG). Five enzymes, Sus, AGPase, AGPGT, SS, and SBE, were characterized to be critical for this process, and the genes encoding these enzymes play important roles in the formation of UDP-glucose (the first step in the conversion of sucrose to starch) and ADPG, transferring ADPG into amyloplasts from cytoplasm, and yielding the end starch (Sullivan et al., 1991; Yang et al., 2004). In this study, differential expression of these rate-limiting enzyme-encoding genes was observed at the middle to late stages of grain development. Haplotypes of *AGPL* have been reported to be associated with TGW, perhaps through transcript level variations affected by a SNP in common wheat (Hou et al., 2017). In this study, *AGPL* co-localized with 1A-2, which involved QTLs for TGW, GW, and GLW (Fig. 2), and the negative allele from KT1-1 could decrease TGW by 1.20 g, accounting for 7.8–15.4% of parental phenotypic variation (Table 2; Supplementary Table S1). Down-regulation of *AGPL* in KT1-1 compared with KT3-5, and a 9 bp deletion (ACTCCGCCG) in the promoter (–1049 to –1057), together with several variants, were identified (Supplementary Fig. S9), indicating that the reduction of transcripts in KT1-1 might be caused by these variants, which is also consistent with the negative effect on GW and TGW. Alleles for *AGPL* and *Sus* have been closely associated with grain weight in wheat, mainly contributed by variation in transcript levels, which has an effect of approximately 3–5 g TGW for *AGPL* and 2–4 g TGW for *Sus* (Jiang et al., 2011; Hou et al., 2017). Overall, 31 of 48 genes in the starch biosynthesis pathway were expressed differentially at different spike development stages, especially the aforementioned rate-limiting enzyme genes (Fig. 5). Our findings therefore indicate that different patterns of expression of these pathway genes might together contribute to final grain weight by affecting starch accumulation throughout grain filling.

Untapped alleles were identified in wild einkorn wheat

Seven QTLs within three regions (7A-1, 7A-2, and 7A-3) on Tm7A were identified for which the wild einkorn wheat (KT1-1) positively contributed to all grain size traits (Table 2; Supplementary Table S6). These data provide evidence that the remnant superior alleles controlling improvement of grain size might have been ignored by preliminary selection for domestication ~10 000 years ago, and some alleles that contribute to a wide adaptation to different environments (such as *Vrn3*) were left in wild einkorn wheat. Thus, we speculate that there may be dominant genes or alleles that enlarge grain size and, further, increase grain weight, in natural wild einkorn wheat, and these could be characterized and assessed for their potential in improving yield.

Conclusion

Understanding the genetic architecture of grain traits is a prerequisite to manipulating grain development and improving crop yield potential. We used high-density genetic mapping and genome-wide QTL mapping, homologous gene mapping, and transcriptome analysis to explore the genetic loci of grain traits in einkorn wheat in various environments. We found that a total of 44 genes that are homologous to QTL regions or in the starch biosynthesis pathway showed differential expression, and 20 of them were mapped on nine QTL regions. Our findings demonstrate that the expression patterns of several functional genes are consistent with the allelic effects of the related QTL. The candidate functional genes associated with grain (sink) size and starch biosynthesis were considered to be important components for grain size and starch accumulation in the developing grains, and, in turn, for grain weight. Furthermore, the phenotypic values of the related traits significantly increased as the numbers of positive QTL alleles accumulated, which could be exploited to fine-tune grain size and weight. We have investigated the complex genetic architecture of grain size on a genome-wide scale, thus elucidating QTLs and their underlying genes through genetic mapping and transcriptional profiling, and thereby allowing us to identify candidate genes for grain size and weight and assist marker-based selection in wheat breeding improvement.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. Frequency distribution of phenotypic data in five environments for six quantitative traits of parents and RILs.

Fig. S2. Principal component analysis revealing a morphometric model for variation in grain morphology in einkorn wheat RIL population.

Fig. S3. Chromosomal 4AL/5AL translocation and 4A pericentric inversion in hexaploid wheat revealed by comparing barley and hexaploid wheat genomes through the mapped SNP markers from einkorn wheat.

Fig. S4. Genome-wide LOD profiles for six investigated traits across five environments.

Fig. S5. Physical lengths of homologous regions corresponding to sixteen QTL regions.

Fig. S6. Polymorphism of *Vrn3* in einkorn wheat.

Fig. S7. Genetic overlaps between grain size related traits in the einkorn wheat RIL population.

Fig. S8. Phenotypic variation affected by positive numbers of alleles for six quantitative traits using average phenotypic data.

Fig. S9. Polymorphism of *AGPL* in einkorn wheat.

Fig. S10. GO and KEGG enrichment analysis of differentially expressed genes identified from four developmental stages of KT1-1 versus KT3-5.

Table S1. Phenotypic performances and distribution parameters and correlation coefficients for six quantitative traits of parents and RILs in five environments.

Table S2. Correlation coefficients between GL, GW, GLW, GA, GC, TGW, and HD in the RIL population in four environments.

Table S3. Probability loadings from principal component analysis of phenotypic data in four environments and mean and BLUP values.

Table S4. Barcode and sequencing information of RAD-seq.

Table S5. The high-density genetic linkage map of einkorn wheat.

Table S6. Individual QTL detected with the CIM method from five environments using the high-density SNP map of einkorn wheat.

Table S7. Candidate genes mapped to QTL regions based on homologous analysis with hexaploid wheat, barley, and rice.

Table S8. Polymorphic markers of functional genes.

Table S9. Phenotypic variation explained by detected QTL estimated using ANOVA results from a simple model and multiple regression analysis, and effects of QTL and environments based on ANOVA.

Table S10. Multiple comparison test of phenotypic variation affected by positive numbers of alleles for six quantitative traits across all five environments.

Table S11. GO enrichment test of differentially expressed genes identified from four developmental stages.

Table S12. KEGG enrichment test of differentially expressed genes identified from four developmental stages.

Table S13. Expression profiles of genes involved in the starch biosynthesis pathway and candidate genes mapped to QTL regions based on homologous analysis with hexaploid wheat, barley, and rice.

Acknowledgements

We thank Chikako Shindo and Tetsuo Sasakuma (Kihara Institute for Biological Research/Graduate School of Integrated Science, Yokohama City University, Japan) for providing experimental materials used in this study. This work was supported by the National Key Research and Development Program of China (2016YFD0101802, 2016YFD0101004) and the National Natural Science Foundation of China (31571643).

Competing interests

The authors declare that they have no competing interests.

Author contributions

AZ and CL conceived and supervised the study; KY, DL, YC, DW, WLY, WY, LY, CZ, SZ, and JS conducted the research and analyzed the data; KY and DL planned and conducted the construction of RAD-seq libraries; KY, DW, WLY, and JS collected phenotypic data; KY analyzed the genotypic and phenotypic data; KY collected samples for transcriptomic analysis; KY, WY, LY, CZ, and SZ contributed to analyses of transcriptomic data; KY, DL, and AZ prepared the manuscript. All authors discussed the results and commented on the manuscript.

References

- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA.** 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376.
- Barchi L, Lanteri S, Portis E, et al.** 2012. A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PLoS ONE* **7**, e43740.
- Bertioli DJ, Cannon SB, Froenicke L, et al.** 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature Genetics* **48**, 438–446.
- Brinton J, Simmonds J, Minter F, Leverington-Waite M, Snape J, Uauy C.** 2017. Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *New Phytologist* **215**, 1026–1038.
- Campbell KG, Bergman CJ, Gualberto DG, Anderson JA, Giroux MJ, Hareland G, Fulcher RG, Sorrells ME, Finney PL.** 1999. Quantitative trait loci associated with kernel traits in a soft × hard wheat cross. *Crop Science* **39**, 1184–1195.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH.** 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171–182.
- Chapman JA, Mascher M, Buluç A, et al.** 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology* **16**, 26.
- Chen Z, Wang B, Dong X, Liu H, Ren L, Chen J, Hauck A, Song W, Lai J.** 2014. An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F₂ maize population. *BMC Genomics* **15**, 433.
- Cheng R, Kong Z, Zhang L, Xie Q, Jia H, Yu D, Huang Y, Ma Z.** 2017. Mapping QTLs controlling kernel dimensions in a wheat inter-varietal RIL mapping population. *Theoretical and Applied Genetics* **130**, 1405–1414.
- Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM.** 2011. Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics* **12**, 4.
- Cui F, Zhang N, Fan XL, et al.** 2017. Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. *Scientific Reports* **7**, 3788.
- Devos KM, Dubcovsky J, Dvořák J, Chinoy CN, Gale MD.** 1995. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theoretical and Applied Genetics* **91**, 282–288.
- Feuillet C, Travella S, Stein N, Albar L, Nublát A, Keller B.** 2003. Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. *Proceedings of the National Academy of Sciences, USA* **100**, 15253–15258.
- Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan JH, Snape JW.** 2010. A genetic framework for grain size and shape variation in wheat. *The Plant Cell* **22**, 1046–1056.
- Georgelis N, Braun EL, Shaw JR, Hannah LC.** 2007. The two AGPase subunits evolve at different rates in angiosperms, yet they are equally sensitive to activity-altering amino acid changes when expressed in bacteria. *The Plant Cell* **19**, 1458–1472.
- Golan G, Oksenberg A, Peleg Z.** 2015. Genetic evidence for differential selection of grain and embryo weight during wheat evolution under domestication. *Journal of Experimental Botany* **66**, 5703–5711.
- He ZH, Rajaram S, Xin ZY, Huang GZ.** 2001. A history of wheat breeding in China. Mexico DF: CIMMYT.

- Hori K, Takehara S, Nankaku N, Sato K, Sasakuma T, Takeda K. 2007. Barley EST markers enhance map saturation and QTL mapping in diploid wheat. *Breeding Science* **57**, 39–45.
- Hou J, Li T, Wang Y, Hao C, Liu H, Zhang X. 2017. ADP-glucose pyrophosphorylase genes, associated with kernel weight, underwent selection during wheat domestication and breeding. *Plant Biotechnology Journal* **15**, 1533–1543.
- Hu Y, Yan C, Hsu CH, Chen QR, Niu K, Komatsoulis GA, Meerzaman D. 2014. OmicCircos: a simple-to-use R package for the circular visualization of multidimensional omics data. *Cancer Informatics* **13**, 13–20.
- Iehisa JC, Ohno R, Kimura T, Enoki H, Nishimura S, Okamoto Y, Nasuda S, Takumi S. 2014. A high-density genetic map with array-based markers facilitates structural and quantitative trait locus analyses of the common wheat genome. *DNA Research* **21**, 555–567.
- International Wheat Genome Sequencing Consortium. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788.
- Jia J, Zhao S, Kong X, et al. 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95.
- Jiang C, Fu X. 2007. GA action: turning on de-DELLA repressing signaling. *Current Opinion in Plant Biology* **10**, 461–465.
- Jiang Q, Hou J, Hao C, Wang L, Ge H, Dong Y, Zhang X. 2011. The wheat (*T. aestivum*) sucrose synthase 2 gene (*TaSus2*) active in endosperm development is associated with yield traits. *Functional & Integrative Genomics* **11**, 49–61.
- Jing HC, Kornukhin D, Kanyuka K, Orford S, Zlatska A, Mitrofanova OP, Koebner R, Hammond-Kosack K. 2007. Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in *Triticum monococcum*. *Journal of Experimental Botany* **58**, 3749–3764.
- Komatsuda T, Maxim P, Senthil N, Mano Y. 2004. High-density AFLP map of nonbrittle rachis 1 (*btr1*) and 2 (*btr2*) genes in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **109**, 986–995.
- Kosambi DD. 1943. The estimation of map distances from recombination values. *Annals of Eugenics* **12**, 172–175.
- Krasileva KV, Vasquez-Gross HA, Howell T, et al. 2017. Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences, USA* **114**, E913–E921.
- Kuchel H, Williams K, Langridge P, Eagles HA, Jefferies SP. 2007. Genetic dissection of grain yield in bread wheat. II. QTL-by-environment interaction. *Theoretical and Applied Genetics* **115**, 1015–1027.
- Kumar A, Mantovani EE, Seetan R, et al. 2016. Dissection of genetic factors underlying wheat kernel shape and size in an elite × nonadapted cross using a high density SNP linkage map. *The Plant Genome* **9**. doi: 10.3835/plantgenome2015.09.0081.
- Lewis S, Faricelli ME, Appendino ML, Valárik M, Dubcovsky J. 2008. The chromosome region including the earliness *per se* locus *Eps-Am1* affects the duration of early developmental phases and spikelet number in diploid wheat. *Journal of Experimental Botany* **59**, 3595–3607.
- Li W, Yang B. 2017. Translational genomics of grain size regulation in wheat. *Theoretical and Applied Genetics* **130**, 1765–1771.
- Maccafferri M, Ricci A, Salvi S, 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**, 648–663.
- Maphosa L, Langridge P, Taylor H, et al. 2014. Genetic control of grain yield and grain physical characteristics in a bread wheat population grown under a range of environmental conditions. *Theoretical and Applied Genetics* **127**, 1607–1624.
- Mascher M, Gundlach H, Himmelbach A, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433.
- Peleg Z, Fahima T, Korol AB, Abbo S, Saranga Y. 2011. Genetic analysis of wheat domestication and evolution under domestication. *Journal of Experimental Botany* **62**, 5051–5061.
- Pfender WF, Saha MC, Johnson EA, Slabaugh MB. 2011. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theoretical and Applied Genetics* **122**, 1467–1480.
- Pourkheirandish M, Hensel G, Kilian B, et al. 2015. Evolution of the grain dispersal system in barley. *Cell* **162**, 527–539.
- Prashant R, Kadoo N, Desale C, Kore P, Dhaliwal HS, Chhuneja P, Gupta V. 2012. Kernel morphometric traits in hexaploid wheat (*Triticum aestivum* L.) are modulated by intricate QTL × QTL and genotype × environment interactions. *Journal of Cereal Science* **56**, 432–439.
- R Core Team. 2016. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> (last accessed June 9, 2019).
- Rasheed A, Xia X, Ogbonnaya F, Mahmood T, Zhang Z, Mujeeb-Kazi A, He Z. 2014. Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. *BMC Plant Biology* **14**, 128.
- Russo MA, Ficco DBM, Laidò G, Marone D, Papa R, Blanco A, Gadaleta A, De Vita P, Mastrangelo AM. 2014. A dense durum wheat × *T. dicoccum* linkage map based on SNP markers for the study of seed morphology. *Molecular Breeding* **34**, 1579–1597.
- Saghai-Marouf MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences, USA* **81**, 8014–8018.
- Saintenac C, Jiang D, Wang S, Akhunov E. 2013. Sequence-based mapping of the polyploid wheat genome. G3: Genes, Genomes, Genetics **3**, 1105–1114.
- Segami S, Kono I, Ando T, Yano M, Kitano H, Miura K, Iwasaki Y. 2012. *Small and round seed 5* gene encodes alpha-tubulin regulating seed cell elongation in rice. *Rice* **5**, 4.
- Shindo C, Sasakuma T, Watanabe N, Noda K. 2002. Two-gene systems of vernalization requirement and narrow-sense earliness in einkorn wheat. *Genome* **45**, 563–569.
- Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai YS, Gill BS, Faris JD. 2006. Molecular characterization of the major wheat domestication gene *Q*. *Genetics* **172**, 547–555.
- Singh K, Ghai M, Garg M, Chhuneja P, Kaur P, Schnurbusch T, Keller B, Dhaliwal HS. 2007. An integrated molecular linkage map of diploid wheat based on a *Triticum boeoticum* × *T. monococcum* RIL population. *Theoretical and Applied Genetics* **115**, 301–312.
- Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D. 2005. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology* **23**, 75–81.
- Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B. 2000. Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). *Proceedings of the National Academy of Sciences, USA* **97**, 13436–13441.
- Sullivan TD, Strelow LI, Illingworth CA, Phillips RL, Nelson OE Jr. 1991. Analysis of maize *brittle-1* alleles and a defective Suppressor-mutator-induced mutable allele. *The Plant Cell* **3**, 1337–1348.
- Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in RNA-seq: a matter of depth. *Genome Research* **21**, 2213–2223.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511–515.
- Tsilo TJ, Hareland GA, Simsek S, Chao S, Anderson JA. 2010. Genome mapping of kernel characteristics in hard red spring wheat breeding lines. *Theoretical and Applied Genetics* **121**, 717–730.
- Tyagi S, Mir RR, Balyan HS, Gupta PK. 2015. Interval mapping and meta-QTL analysis of grain traits in common wheat (*Triticum aestivum* L.). *Euphytica* **201**, 367–380.
- Van Ooijen JW, Voorrips RE. 2004. JoinMap Version 4.0, software for the calculation of genetic linkage maps. Wageningen: Kyazma BV.
- Voorrips RE. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**, 77–78.
- Wang N, Fang L, Xin H, Wang L, Li S. 2012a. Construction of a high-density genetic map for grape using next generation restriction-site associated DNA sequencing. *BMC Plant Biology* **12**, 148.
- Wang RX, Hai L, Zhang XY, You GX, Yan CS, Xiao SH. 2009. QTL mapping for grain filling rate and yield-related traits in RILs of the Chinese winter

- wheat population Heshangmai × Yu8679. *Theoretical and Applied Genetics* **118**, 313–325.
- Wang S, Basten CJ, Zeng Z-B.** 2012b. Windows QTL Cartographer 2.5. Raleigh: Department of Statistics, North Carolina State University. <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm> (last accessed June 9, 2019).
- Williams K, Sorrells ME.** 2014. Three-dimensional seed size and shape QTL in hexaploid wheat (*Triticum aestivum* L.) populations. *Crop Science* **54**, 98–110.
- Wu QH, Chen YX, Zhou SH, et al.** 2015. High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 × Beinong6. *PLoS ONE* **10**, e0118144.
- Wu W, Li C, Ma B, Shah F, Liu Y, Liao Y.** 2014. Genetic progress in wheat yield and associated traits in China since 1945 and future prospects. *Euphytica* **196**, 155–168.
- Wu Y, Bhat PR, Close TJ, Lonardi S.** 2008. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genetics* **4**, e1000212.
- Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q.** 2010. Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proceedings of the National Academy of Sciences, USA* **107**, 10578–10583.
- Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J.** 2004. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* **303**, 1640–1644.
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J.** 2003. Positional cloning of the wheat vernalization gene *VRN1*. *Proceedings of the National Academy of Sciences, USA* **100**, 6263–6268.
- Yang J, Zhang J, Wang Z, Xu G, Zhu Q.** 2004. Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. *Plant Physiology* **135**, 1621–1629.
- Yu K, Liu D, Wu W, et al.** 2017. Development of an integrated linkage map of einkorn wheat and its application for QTL mapping and genome sequence anchoring. *Theoretical and Applied Genetics* **130**, 53–70.
- Zaharieva M, Monneveux P.** 2014. Cultivated einkorn wheat (*Triticum monococcum* L. subsp. *monococcum*): the long life of a founder crop of agriculture. *Genetic Resources and Crop Evolution* **61**, 677–706.
- Zhang T, Hu Y, Jiang W, et al.** 2015. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology* **33**, 531–537.
- Zhang Y, Luo G, Liu D, Wang D, Yang W, Sun J, Zhang A, Zhan K.** 2015. Genome-, transcriptome- and proteome-wide analyses of the gliadin gene families in *Triticum urartu*. *PLoS ONE* **10**, e0131559.
- Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, Hong W, Zhang X.** 2013. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biology* **13**, 141.