ORIGINAL ARTICLE

Genotyping-by-sequencing and genome-wide association study reveal genetic diversity and loci controlling agronomic traits in triticale

Dong Cao[1,2,3,4] · Dongxia Wang[5] · Shiming Li[1,2] · Yun Li[1,2,3] · Ming Hao[6] · Baolong Liu[1,2,3,4]

Abstract

Key message **The genetic diversity and loci underlying agronomic traits were analyzed by the reads coverage and genome-wide association study based genotyping-by-sequencing in a diverse population consisting of 199 accessions.**

Abstract Triticale (× *Triticosecale* Wittmack) is an economically important grain forage and energy crop planted worldwide for its high biomass. Little is known about the genetic diversity and loci underlying agronomic traits in triticale. We performed genotyping-by-sequencing of 199 cultivars and mapped reads to the A, B, D, and R genomes for karyotype analysis. These cultivars could mostly be grouped into five types. Some chromosome abnormalities occurred with high frequency, such as 2D (2R) substitution, deletion of the long arm of chromosome 2D or the short arm of 5R, and translocation of the long arms of 7D/7A, the short arms of 6D/6A, or the long arms of 1D/1A. We chose only widely planted hexaploid triticale cultivars (153) for genome-wide association study. These cultivars could be divided into nine distinct groups, and the linkage disequilibrium decay was 25.4 kb in this population. We identified 253 significant marker-trait associations (MTAs) on 20 chromosomes, except 7R. Twenty-one reliable MTAs were identified repeatedly over two environments. We predicted 16 putative candidate genes involved in plant growth and development using the genome sequences of wheat and rye. These results provide a basis for understanding the genetic mechanisms of agronomic traits and will benefit the breeding of improved hexaploid triticale.

Communicated by Thomas Lubberstedt.

✉ Ming Hao
haomingluo@foxmail.com

✉ Baolong Liu
blliu@nwipb.cas.cn

1. Qinghai Province Key Laboratory of Crop Molecular Breeding, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, Qinghai 810008, People's Republic of China

2. Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, Qinghai 810008, People's Republic of China

3. The Innovative Academy of Seed Design, Chinese Academy of Sciences, Xining 810008, Qinghai, China

4. University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

5. Department of Agriculture and Forestry, College of Agriculture and Animal Husbandry, Qinghai University, Qinghai, Xining 810016, People's Republic of China

6. Triticeae Research Institute, Sichuan Agricultural University, Chengdu 611130, People's Republic of China

Abbreviations

GBS	Genotyping-by-sequencing
GWAS	Genome-wide association study
MTAs	Marker-trait associations
PH	Plant height
UIL	Uppermost internode length
SL	Spike length
SNS	Spikelet number per spike
GNS	Grain number per spike
SSP	Seed setting percentage
TGW	Thousand-grain weight
GL	Grain length
GW	Grain width
GA	Grain area

Introduction

Triticale (× *Triticosecale* Wittmack) is a man-made, amphiploid species possessing high yield and abiotic stress resistance characteristics (Blum 2014). It is usually used for food production and animal feed as a grain-forage crop (McGoverin et al. 2011; Zhu 2018). High biomass

production also makes it a potential energy crop for industrial purposes, as bioenergy and biofuel (McGoverin et al. 2011). The planting area of triticale has continued to grow as demand increases, with major producers including Poland, Germany, Belarus, France, Russia, and China (FAOSTAT 2017).

The first triticale was obtained by distant hybridization between common wheat (*Triticum aestivum* L.) and rye (*Secale cereale*), and was octoploid ($2n = 8x = 56$, AABBDDRR) (Mergoum and Gómez-Macpherson 2004). Hexaploid triticales ($2n = 6x = 42$, AABBRR) were also created using tetraploid wheat (*Triticum turgidum* L.) and rye, and tetraploid ($2n = 28 = $ AARR) and decaploid ($2n = 70 = $ ABBDDRRRR) triticales likewise exist. Although the National Plant Germplasm System of America has 1971 triticale resources deposited, the genetic backgrounds of these cultivars remain unknown. Distant hybridization causes unstable ploidy levels resulting in tetraploid, octoploid, and decaploid triticales because of chromosome incompatibility. Compared with other triticale, hexaploid triticale show relatively superior vigor and reproductive stability and have become the main plant types in worldwide (Cheng and Murata 2002; Fox et al. 1990; Lukaszewski and Gustafson 1987). Genetic diversity of hexaploid triticale were evaluated based on few simple sequence repeats (SSR) and diversity array technology (DArT) markers from wheat and rye (Costa et al. 2007; Kuleung et al. 2006; Niedziela et al. 2016; Tams et al. 2004). Several genetic maps have been constructed for hexaploid triticale based on DArT, single-nucleotide polymorphism (SNP)-DArT, SSR, and amplified fragment length polymorphism (AFLP) markers in different populations (Alheit et al. 2011; Tyrka et al. 2011, 2015, 2018). These maps have been used to identify loci related to plant height (PH), spike length (SL), spikelet number per spike (SNS), grain number per spike (GNS), seed setting percentage (SSP), thousand-grain weight (TGW) and other traits in hexaploid triticale (Alheit et al. 2014; Liu et al. 2014, 2016; Wajdzik et al. 2019; Würschum et al. 2014). However, the molecular markers used in these studies were limited, and the plant materials were restricted to a few samples. Moreover, all previous research was carried out without reference genome information.

Here, we collected 199 triticale cultivars originating from different countries from the National Plant Germplasm System of America and our own collections for evaluating the genetic variation of triticale based on genotyping-by-sequencing (GBS) technology (Akram et al. 2021; Sakiroglu et al. 2017; Torkamaneh et al. 2021). Wheat (*Triticum* sp.) (v1.0, International Wheat Genome Sequencing Consortium, IWGSC) and rye (*Secale* sp.) genome sequences were used as reference genomes (Rabanus-Wallace et al. 2021) for mapping reads and SNP calling. We then carried out genome-wide association study (GWAS) to identify the loci and candidate genes controlling agronomic traits in hexaploid triticale.

Materials and methods

Plant material

Plant material consisted of 199 triticale accessions, of which 191 were obtained from the SDA-ARS National Small Grains Collection (http://www.ars-grin.gov/) and 19 were provided by the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences. The 191 triticale accessions were classified as × *Triticosecale* spp. according to information provided by the USDA-ARS website. The accessions were derived from China (37), United States (26), Mexico (13), Canada (11), Russia (11), Japan (10), Sweden (10), Poland (9), South Africa (9), India (8), Bulgaria (8), France (7), Spain (7), Australia (6), Germany (6), Hungary (5), Brazil (4), Ukraine (4), Azerbaijan (2), Armenia (1), Ecuador (1), England (1), Italy (1), Morocco (1), and Romania (1).

Genotyping-by-sequencing

Leaf tissues of 199 triticale accessions from single plant were collected at seedling stage. Genomic DNA was extracted from leaves using the cetyltrimethylammonium bromide (CTAB) method (Yan et al. 2002) and used to construct libraries for genotyping-by-sequencing (GBS) analysis. GBS libraries were prepared using protocols described by Poland et al. (2012). Libraries were sequenced using Illumina paired-end sequencing technology with read lengths of 150 bp on an Illumina HiSeq X instrument by Novogene Bioinformatics Institute (Beijing, China). Raw reads were subjected to a set of quality control procedures using FastQC v0.11.5 and Trimmomatic-0.38. The sequence reads were stored in the National Center for Biotechnology Information (NCBI) SRA database with the accession number SUB10900113. The reference genome of the wheat variety Chinese Spring (RefSeq V1.0) was selected, and publication access to the rye DNA assembly and annotation was kindly provided by Drs. N. Stein and M. Wallace of the Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany (Rabanus-Wallace et al. 2021). After excluding reads with low quality bases (quality score < 20) and length less than 50 bp, clean reads were mapped against the wheat and rye genomes using the Burrows-Wheeler Aligner (BWA) software with the BWA-MEM algorithm (Li and Durbin, 2009) and then subjected to single-nucleotide polymorphism (SNP) calling using the GATK software. SNP filtering was performed using the following parameters: reads with minor allele frequency (MAF) less than 5%, missing rate more than 20%, QualByDepth

(QD) less than 2.0, RMSMappingQuality (MQ) < 30.0, FisherStrand (FS) > 60.0, MappingQualityRankSumTest (MQRankSum) < − 25, ReadPosRankSumTest (ReadPosRankSum) < − 8.0, DP less than 150 and more than 2000, and Quality (QUAL) less than 30 were filtered out. SNPs data was uploaded to figshare (https://doi.org/10.6084/m9.figshare.17204117). A total of 434,304 reads were obtained for genetic diversity analysis. Read coverage on each chromosome was counted, and distribution figures were drawn. These accessions could be divided into five types according to SNP distribution and fluorescence in situ hybridization (FISH) analysis: octaploid triticale (AABBDDRR), hexaploid triticale (AABBRR), hexaploid wheat (AABBDD), tetraploid wheat (AABB), and rye (RR). Detailed information for these accessions is described in Table S1.

FISH analyses

Seed germination and root tip collection and processing were as described by Zhao et al. (2018). Slides were prepared as described by Komuro et al. (2013). Three oligonucleotide (oligo) probes with 6-FAM or Tamra label at the 5′-end, (AAC)5 (Dennis et al. 1980), oligo-pSc119.2 and oligo-pTa535 (Tang et al., 2014), were synthesized by the TsingKe Biological Technology Company (Chengdu, China) and used to identify chromosome constitution. oligo-pSc119.2 was used to detect B- and R-genome chromosomes, and oligo-pTa535 was used to detect A- and D-genome chromosomes. The FISH procedure was conducted according to methods described by Zhao et al. (2018). Photomicrographs of chromosome observations were created and documented using an Olympus BX-63 microscope (Tokyo, Japan) coupled to a Photometric SenSys Olympus DP80 CCD camera. Karyotype analysis was carried out on 5 accessions selected according to read coverage of GBS results.

Field trial and phenotyping

A bulk of 153 accessions was planted using a randomized complete block design with three replicates at two sites in Qinghai, China, namely Chengbei (CB, 101.6° E, 36.7° N) and Huang zhong (HZ, 101.5° E, 36.5° N), over 2 years (2019–2020). Each replicate row was 2 m long, with 20 cm row-to-row distance, and 50 seeds per sow. All accessions were sown in march each year. Nitrogen (urea) and phosphate (ammonium phosphate) fertilizers were applied with a rate of 120 kg/ha and 300 kg/ha, respectively. Field management, disease and pest control were adopted according to the common practices as per need. Ten representative and single plants were selected for measuring ten agronomic traits in three environments: E1 (CB/2019), E2 (CB/2020), and E3 (HZ/2020). Plant height (PH), uppermost internode length (UIL), spike length (SL), and spikelet number per

spike (SNS) were measured at the maturity stage, and grain number per spike (GNS), seed setting percentage (SSP), thousand-grain weight (TGW), grain length (GL), grain width (GW), and grain area (GA) were measured after harvesting. GL, GW, and GA were measured directly using a high-flux Marvin seed analyzer (GTA Sensorik GmbH, Neubrandenburg, Germany). Phenotypic data for the 10 agronomic traits were analyzed using Microsoft Excel 2010 and SPSS 11.5 software. Broad-sense heritability (H^2) per trait was estimated by analysis of variance (ANOVA) using the R software package "lme4" (Bates et al. 2015; Singh et al. 1993). Pearson's correlation coefficients between pairs of agronomic traits were determined using the R software package "corrplot".

Population genetic analyses

SNPs for distances less than 10 bp were filtered out, and 392,915 SNPs from 153 hexaploid triticale accessions were used for further GWAS analysis. SNP distribution, principal component analysis (PCA), reconstruction of a neighbor-joining tree, linkage disequilibrium (LD) decay, and population structure analysis were carried out using the valid SNPs. SNP density distribution was depicted using the R package with CMplotmodule (http://ftp.hu.debian.org/pub/CRAN/web/packages/CMplot/index.html). Pairwise LD and r^2 between SNPs were calculated using PLINK v1.07 software (Purcell et al. 2007). Population structure analysis was performed on 153 hexaploid triticale accessions using ADMIXTURE software (Alexander et al. 2009). The ADMIXTURE program was run with K values ranging from 2 to 20. The best K value was estimated to confirm the number of groups based on cross-validation (CV) and log-likelihood estimates. MEGA5 software with the neighbor-joining method was used to reconstruct the phylogenetic tree (Tamura et al. 2011).

GWAS

The R package of the Genome Association and Prediction Integrated Tool (GAPIT) was employed for GWAS using the Fixed and random model Circulating Probability Unification (FarmCPU) model. SNPs and traits were significantly linked if the P value threshold was less than 0.05/N, where N was the total number of SNP markers used for association analysis (Duggal et al. 2008; Li et al. 2012; Yang et al. 2014). The significant P value threshold was $1.27E^{-07}$ ($-\log 10(P) = 6.90$) for screening the valid SNPs.

Putative candidate gene analysis

Genes significantly associated with MTAs were predicted according to a previously described method (Rahimi et al.
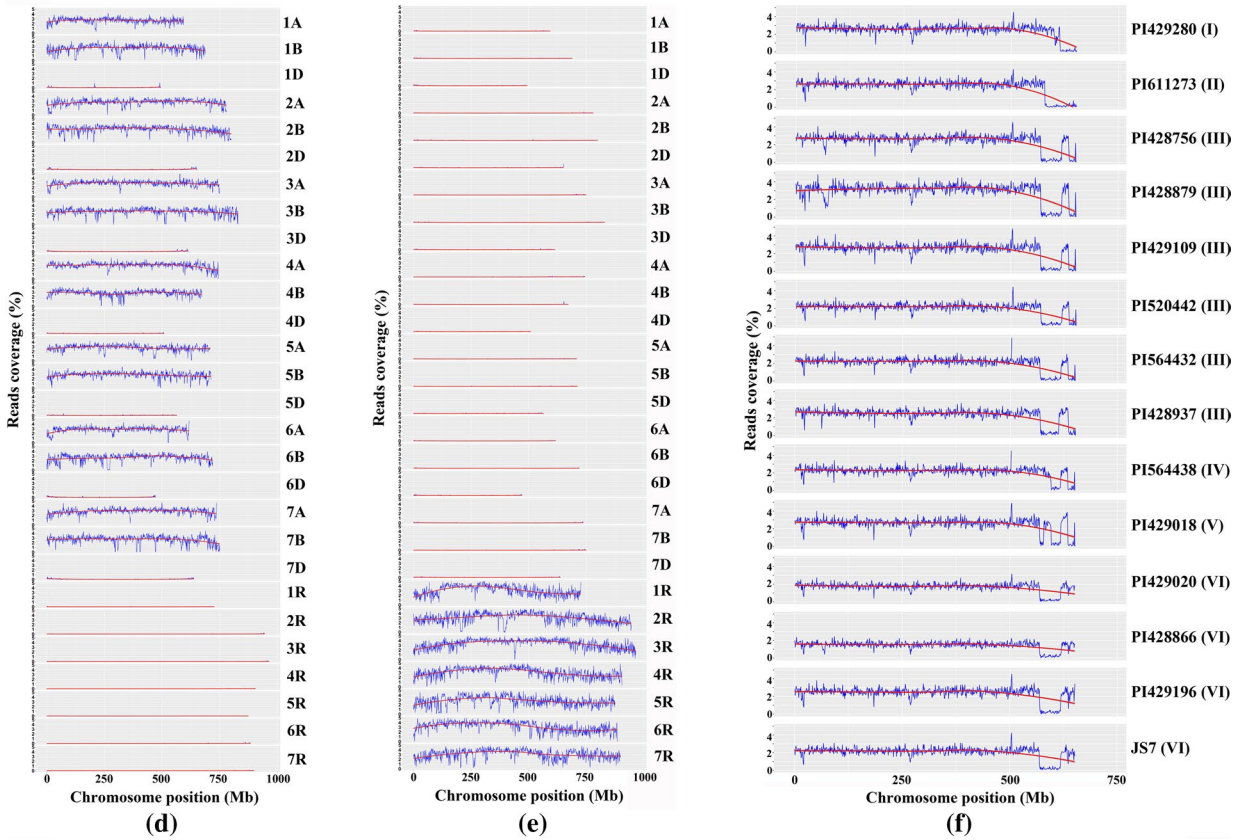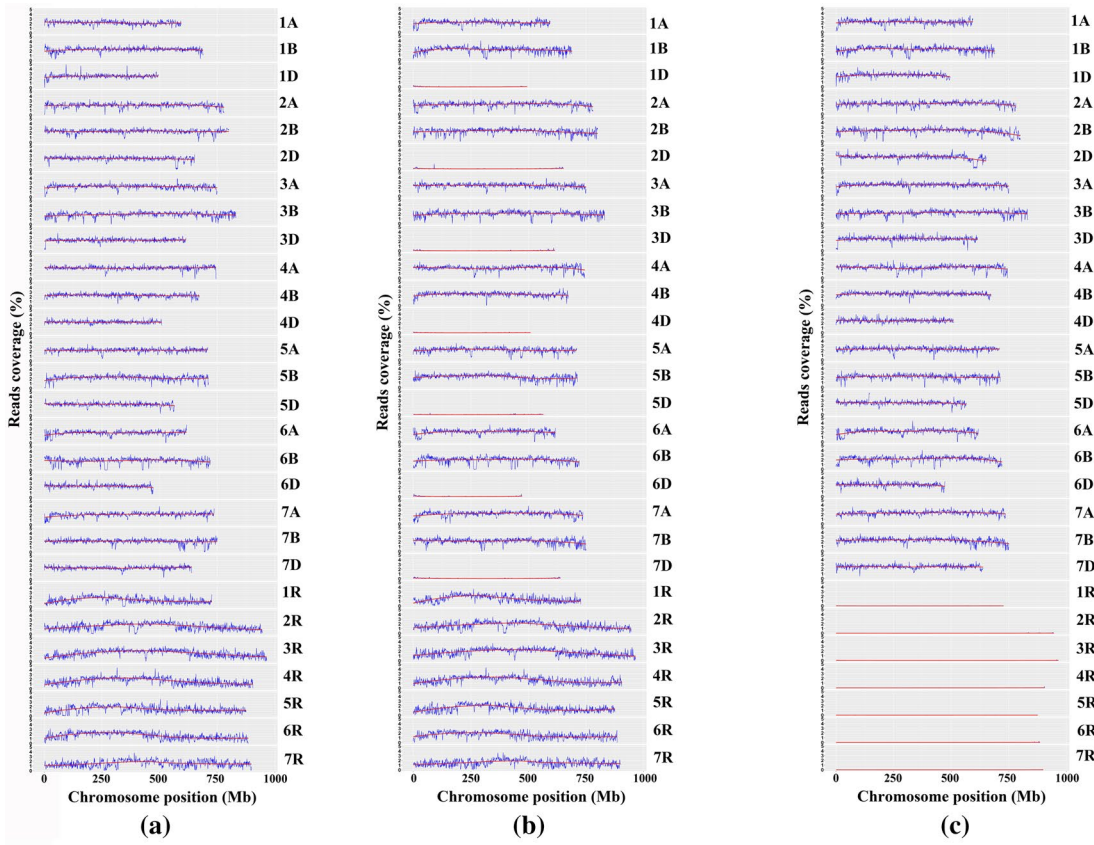
◄**Fig. 1** Read coverage in five ploidy species. **a** Octaploid triticale (AABBDDRR), **b** hexaploid triticale (AABBRR), **c** hexaploid wheat (AABBDD), **d** tetraploid wheat (AABB), **e** rye (RR). **f** Partial deletion of the long arm of chromosome 2D in 14 accessions based on read coverage. Deleted segments are divided into type I to VI based on length

2019). The functions of the predicted genes were annotated using BLASTALL search.

Results

Assessment of genetic diversity by GBS

We genotyped 199 *Triticosecale* spp. cultivars from the National Plant Germplasm System of America and our own collections using the GBS method. Triticale usually has an unstable ploidy. We therefore developed a method to analyze ploidy levels using reads mapped to chromosomes A, B, D, and R. The 199 accessions could be easily classified into five ploidy species based on reads coverage: 22 octaploid triticale (AABBDDRR), 153 hexaploid triticale (AABBRR), 20 hexaploid wheat (AABBDD), three tetraploid wheat (AABB), and one rye (RR) (Fig. 1a–e; Supplementary Table S1). Cytological analyses further supported the results (Fig. S1). The read mapping also revealed the deletion and addition of whole chromosomes or chromosome segments. The whole 2R chromosome was frequently lost, which we found in eight cultivars of octaploid triticale and 38 cultivars of hexaploid triticale (Supplementary Table S1). Chromosome 2D was added in 38 hexaploid triticales, with deletion of the 2R chromosome except in one cultivar (Supplementary Table S1). The 2D chromosome was also added in five hexaploid triticale cultivars, without deletion of chromosome 2R (Supplementary Table S1). Chromosome 6A was replaced by chromosome 6D in two hexaploid triticales (Supplementary Table S1). Although 20 cultivars were annotated as hexaploid wheat, one cultivar carried an additional chromosome 2R and one cultivar had chromosomes 6D and 1D replaced by 6R and 1R, indicating that these two cultivars were derived from octaploid triticale (Supplementary Table S1).

Some chromosome segments were also abnormal in these cultivars. For example, the short arm of 5R was partially lost in seven cultivars of octaploid triticale, with a similar segment length in all cultivars (Fig. S2). Fourteen cultivars lost part of the long arm of chromosome 2D, and the lost segment was shorter than the 5RS chromosome (Fig. 1f). The length of the deleted segments was different in the 14 cultivars, which could be divided into six types: two octaploid triticale, seven hexaploid triticale, and five hexaploid wheat (Fig. 1f; Supplementary Table S1). Obvious chromosomal

translocations were observed on the long arm of 7D/7A, the short arm of 6D/6A, and the long arm of 1D/1A in three, two, and two cultivars, respectively (Fig. S3). Moreover, four events of segment insertion, and three events of segment deletion occurred in ony one cultivar (Fig. S4). The location of the segment insertion remains unknown.

Single nucleotide polymorphism and phenotypic variation of hexaploid triticales

Because hexaploid triticale is also the main cultivated type (Lukaszewski and Gustafson 1987; Cheng and Murata 2002; Fox et al. 1990), we chose 153 hexaploid triticale cultivars for further SNP calling, phenotype evaluation, and GWAS analysis. A total of 434,304 reliable SNP markers were obtained in 153 hexaploid triticale cultivars. These SNPs were not evenly distributed among sub-genomes A, B, and R and 28.13 SNPs/Mb were detected for the entire genome (Fig. S5; Supplementary Table S2). Sub-genome B harbored the most SNPs (212,218) with the highest SNP density (24.56 kb/SNP), followed by A (187,518, 27.92 kb/SNP) and R (34,568, 3001.60 kb/SNP) (Supplementary Table S2). The ratio of the number of SNPs in the B or A genome to those in the R genome was 6.14 and 5.42, respectively. Sub-genome R harbored the least SNPs and the lowest SNP density compared with A and B, and the number of SNPs in the A genome was only slightly lower than that in the B genome. Chromosome 7A had the most SNPs (36,228) with the highest SNP density (20.32 kb/SNP), and 2R had the least SNPs (44) with the lowest SNP density (20,015.53 kb/SNP) (Supplementary Table S2). To further reveal the genetic relationships of the genotypes, we estimated the optimal number of genetic clusters determined (K), ranging from 1 to 20 (Fig. 2a). The cross-validation (CV) error was minimized at $k = 9$ (Fig. 2a). Consistent with the neighbor-joining clustering result (Fig. 2b), the 153 hexaploid triticale genotypes could be classified into nine genetically distinct population groups (POP I–IX) (Fig. 2b, 3c). Through the prediction methods using the genetic distance between the parental lines couldn't predict the hybrid performance of inter-pool hybrids in plant breeding programs reliably (Melchinger 1999), the amazing diversity pattern may serve as a reference to systematically establish test-crosses and determine general combining ability estimates based on field data. The average decay distance of linkage disequilibrium (LD) was about 25.4 kb at the threshold of $r^2 = 0.45$ across all chromosomes (Fig. 2d).

Phenotypes of 153 hexaploid triticale were investigated in three environments to analyze the distribution, variation, and correlation of 10 agronomic traits. All agronomic traits showed a normal distribution, and a linear correlation existed in observed and expected values (Fig. S6, S7). The maximum values of PH, SL, uppermost internode length

Fig. 2 Population structure of 153 hexaploid triticale accessions. **a** Estimation of population using CV error with K ranging from one to 20. **b** Neighbor-joining tree analysis of 153 hexaploid triticale accessions. **c** Population structure based on $k = 9$. **d** Decay of linkage disequilibrium (LD) in genomes of 153 hexaploid triticale accessions
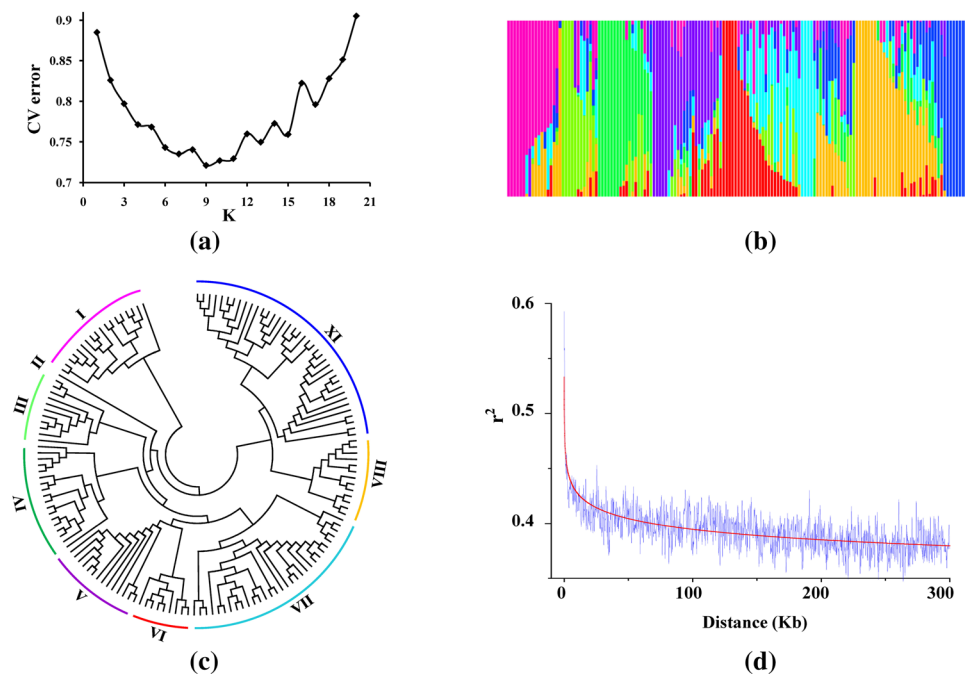
(UIL), SNS, GNS, SSP, TGW, grain area (GA), grain width (GW), and grain length (GL) were 200.60 cm, 15.72 cm, 45.50 cm, 31.52, 86.82, 82.47%, 57.95 g, 28.87 mm^2, 3.70 mm, and 9.58 cm, respectively, while the minimum values were 95.97 cm, 8.43 cm, 16.10 cm, 16.32, 25.05, 46.68%, 27.27 g, 17.70 mm^2, 2.66 mm, and 6.86 cm, respectively (Supplementary Table S3). The maximum values of all agronomic traits except for GA, GW, and GL were more than twice those of the matching minimum values (Supplementary Table S3). The GNS showed the highest coefficient of variation (CV) (20.02%), while GW had the lowest CV (5.80%) (Supplementary Table S3). Broad-sense heritability (H^2) was calculated for the 10 agronomic traits, all traits showed high H^2 except for SSP ($H^2 = 0.37$) and UIL ($H^2 = 0.47$) (Supplementary Table S3). This suggested that genetic components caused the variation in these accessions.

GWAS of agronomic traits in hexaploid triticale

GWAS is an effective method that uses natural populations to identify loci controlling complex traits. Grain yield is the most important index for triticale breeding, and the 10 agronomic traits observed in this research are closely associated with grain yield. In association analysis, we detected 253 marker-trait associations (MTAs) in three environments using the FarmCPU model at a significance value of $-\log_{10}(P) > 6.90$ (Supplementary Table S4). The number of MTAs for PH, SL, UIL, SNS, GNS, SSP, TGW, GA, GW, and GL were 85, 26, 9, 52, 6, 8, 16, 2, 2, and 47, respectively. The number of MTAs reflects the complexity of the traits to a certain extent. These MTAs were distributed

across 20 chromosomes, except for 7R, and explained 0.09–31.46% of the phenotypic variance. Of the MTAs, 105 were located on sub-genome A, 93 on sub-genome B, and 55 on sub-genome R. Chromosome 2A had the most MTAs (30 MTAs), while the least MTAs were present on chromosome 5R (five MTAs). We identified 21 reliable MTAs over two environments and three reliable MTAs over three environments (Fig. 3). Three loci were pleiotropic. 3Brs508522077 was associated with SL and SNS, 6Ars3762529 was associated with PH and UIL, and 1rs122212247 was associated with SNS and PH (Supplementary Table S4).

High quality wheat and rye references are useful resources for identification of candidate genes that underlie marker loci associated with agronomic traits in hexaploid triticale. We predicted genes associated significantly with MTAs for the annotation of these MTAs, and the functions of the predicted genes were annotated using a BLASTALL search. A total of 205 genes associated with 94 significant markers were predicted in the wheat and rye genomes (Supplementary Table S4). The functions of 16 genes were associated with the related traits (Table 1). Five genes were related to PH, and these encoded CASP-like protein, methyl jasmonate-inducible lipoxygenase 2, sugar transporter SWEET14-like, class I glutamine amidotransferase, and auxin transport protein BIG. The sugar transporter SWEET14-like was also associated with UIL. One stable MTA controlling SL was linked to transcripts TraesCS7A02G058900.1 and TraesCS7A02G059000.1, which were annotated as brassinosteroid LRR receptor kinase (Chono et al. 2003; Dockter et al. 2014) and DELLA protein DWARF8-like (Cassani et al. 2009), respectively. Two stable MTAs for
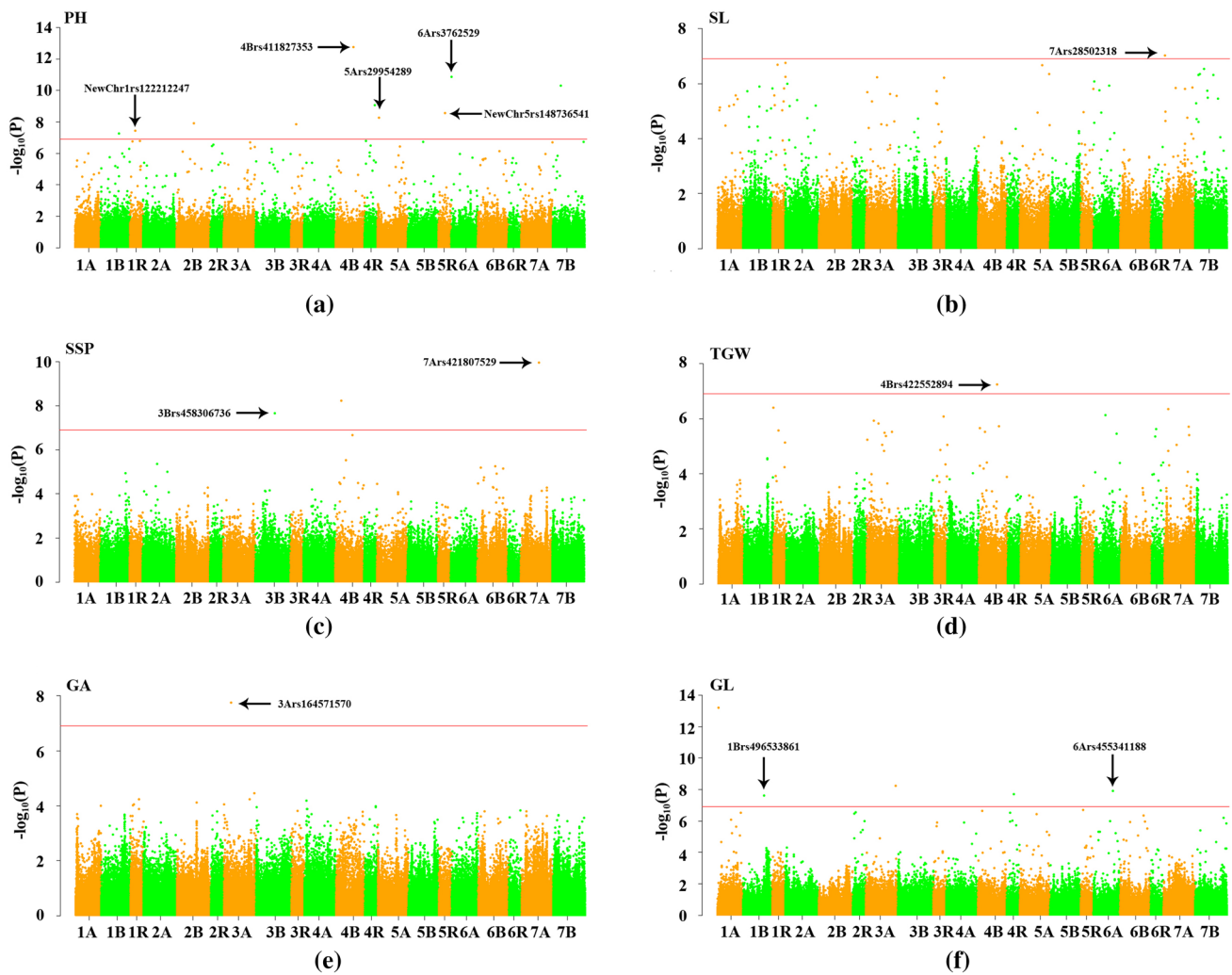
Fig. 3 Manhattan plots of MTAs for **a** PH, **b** SL, **c** SSP, **d** TGW, **e** GA, and **f** GL. The red line indicates the significance threshold (− log10 $P = 6.90$). Stable MTAs are labeled, and putative candidate genes corresponding to these stable MTAs are listed in Table 1

SSP corresponded to transcripts TraesCS3B02G286500.1 and TraesCS7A02G301900.1, annotated as myb-related protein Hv33-like and DNA-directed RNA polymerase II subunit RPB1, respectively (Li et al. 2019, 2020). The gene encoding GS3/putative transmembrane protein is a major gene for grain length and weight in rice (Fan et al. 2006). TraesCS4B02G196100.1 associated with TGW was annotated as a transmembrane protein and was homologous to the gene responsible for grain length and weight in rice (Fan et al. 2006). Ribosomal protein S6 kinase S6K plays an important role in cell proliferation (Henriques et al. 2013), and genes underlying cell proliferation can effectively increase grain size and yield (Brinton et al. 2017; Wang et al. 2012). TraesCS3A02G162200.1 and TraesCS3A02G162300.1 associated with GA encode ribosomal protein S6 kinase, which can promote cell proliferation and increase grain size and yield effectively. Two markers associated with GL were predicted as TraesCS1B02G285400.1,

TraesCS1B02G285700.1, and TraesCS6A02G243900, which encode glucan endo-1,3-beta-glucosidase, GTPase, and xylan 1,4-beta-xylosidase, respectively.

Discussion

Read mapping is a good method for karyotype analysis. Cytological analysis based on microscopy has been used to evaluate chromosome karyotypes for hundreds of years. It requires special cells with mitochysis and special skills to crush cells without destroying the chromosomes. Small chromosomes are difficult to observe and analyze. In this study, we developed a new method to evaluate chromosome morphology based on read mapping without a microscope. It is very convenient to know which chromosomes exist in a sample and to accurately know the length of segments deleted or retained in the chromosomes. Based on read

Table 1 Potential candidate genes linked to significant MTAs

Trait	Environment	SNP	Region	Chromosome	Candidate genes	Annotation	References
PH	E1/E3	4Brs411827353	Intergenic	4B	TraesCS4B02G189300.1	CASP-like protein	Bai et al. (2013)); Li et al. (2011); Yang et al. (2015)
PH	E1/E2/E3	5Ars29954289	Intergenic	5A	TraesCS5A02G032800.1	Methyljasmonate-inducible lipoxygenase 2	Huber et al. (2005); Li et al. (2018)
PH/UIL	E1/E2/E3	6Ars3762529	Intergenic	6A	TraesCS6A02G009000.1 TraesCS6A02G009100.1	Sugar transporter SWEET14-like	Kanno et al. (2016); Sun (2008); Wu et al. (2018)
PH	E2/E3	New-Chr1rs122212247	Upstream	1R	SECCE1Rv1G0059300.1	Class I glutamine amidotransferase	Lou et al. (2021)
PH	E1/E2/E3	New-Chr5rs148736541	Intergenic	5R	SECCE5Rv1G0369160.1	Auxin transport protein	Besseau et al. (2007)
SL	E1/E2	7Ars28502318	Intergenic	7A	TraesCS7A02G058900.1	Brassinosteroid LRR receptor kinase BRI1	Chono et al. (2003); Dockter et al. (2014)
					TraesCS7A02G059000.1	DELLA protein DWARF8-like	Cassani et al. (2009)
SSP	E1/E2	3Brs458306736	Intergenic	3B	TraesCS3B02G286500.1	myb-related protein Hv33-like	Li et al. (2020)
SSP	E1/E3	7Ars421807529	Intergenic	7A	TraesCS7A02G301900.1	DNA-directed RNA polymerase II subunit RPB1	Li et al. (2019)
TGW	E1/E2	4Brs422552894	Intergenic	4B	TraesCS4B02G196100.1	Transmembrane protein	Fan et al. (2006)
GA	E1/E2	3Ars164571570	Intergenic	3A	TraesCS3A02G162200.1 TraesCS3A02G162300.1	Ribosomal protein S6 kinase	Brinton et al. (2017); Wang et al. (2012)
GL	E2/E3	1Brs496533861	Intergenic	1B	TraesCS1B02G285400.1	Glucan endo-1,3-beta-glucosidase	Ayaad et al. (2021)
					TraesCS1B02G285700.1	GTPase	Zhang et al. (2019)
GL	E2/E3	6Ars455341188	Intergenic	6A	TraesCS6A02G243900.1	Xylan 1,4-beta-xylosidase	Yang et al. (2016)

mapping, 199 accessions from various countries were evaluated for their cytological characters. This should represent the relative genetic variation in triticales across the world. We identified five types of materials among these accessions: octaploid triticale (AABBDDRR), hexaploid triticale (AABBRR), hexaploid wheat (AABBDD), tetraploid wheat (AABB), and rye (RR), although all of these materials were labeled as triticale in the National Plant Germplasm System of America. Most (153) were hexaploid triticale. Read mapping could also be used to identify the abnormal behaviors of some chromosomes. For example, chromosome 2R was deleted, and chromosome 2D was retained in hexaploid triticale, which was also observed in previous research (Dou et al. 2006; Hao et al. 2013). Small segment changes could also be identified using this method. Six types of deleted segments were found in the lost segment on the long arm of chromosome 2D. Because the lost segments were very short, it was very hard to identify the differences between them in previous studies. Read mapping also allowed us to

easily obtain the exact length of the deletions. Read mapping represents a convenient and fast method for analyzing chromosome karyotypes.

Previous researchers thought that the chaos of remote hybridization and the incompatibility of the D and R chromosomes caused abnormal chromosome behavior (Dou et al. 2006), and the sequence modification and partial elimination of rye parental genome are common in triticale (Ma and Gustafson 2008; Khalil et al. 2015). This included chromosome 2D being replaced by 2R, and partial deletion of the short arm of chromosome 5R. Partial chromosomal translocations were also observed on the long arm of 7D/7A, the short arm of 6D/6A, and the long arm of 1D/1A, which was not reported in previous studies. The biological functions of these abnormal chromosome behaviors should be studied further. The 199 accessions could be grouped into five types based on the numbers of A, B, D, and R chromosomes. Additional chromosomes revealed the origin of the materials. The sample numbers of tetraploid wheat (AABB) and

rye (RR) were small, and the chromosome behaviors were simple. Abnormal chromosome behavior existed in octaploid triticale (AABBDDRR), hexaploid triticale (AABBRR), and hexaploid wheat (AABBDD). Two hexaploid wheat accessions contained an additional chromosome R, which meant these two cultivars were most likely derived from the octaploid triticale. The 52 hexaploid triticale with a D chromosome or segment should also be derived from the octaploid triticale. Another possibility was that the cross of the hexaploid triticale and common wheat at some stages resulting in the introgression of D chromosomes or segments to hexaploid triticale. Although there were only 22 octaploid triticale, 77 of the 199 cultivars were derived from the octaploid triticales. Because both D and R chromosomes could be lost in future generations of octaploid triticale, it is hard to determine the origin of the other cultivars that do not carry a D genome or R genome. Five hexaploid wheat accessions did not carry any R genome, but these cultivars lost a segment on the long arm of chromosome 2D. It is highly possible that these cultivars were derived from the octaploid triticale. This kind of hexaploid wheat should carry a different structure, which could be used to improve the genetic variation of common wheat.

Karyotype analysis based on read mapping gave us a chance to choose hexaploid triticale for further the population structure and GWAS analysis, which should improve the specificity of the results without the interference of other species. Previous researches had studied the genetic diversity of hexaploid triticale using the limited markers from wheat and rye based on SSR and DArT technologies (Costa et al. 2007; Kuleung et al. 2006; Niedziela et al. 2016; Tams et al. 2004). Based on GBS technology, we obtained 434,304 reliable SNPs, far more than in previous studies (Alheit et al. 2011; Costa et al. 2007; Kuleung et al. 2006; Niedziela et al. 2016; Tams et al. 2004; Tyrka et al. 2011; 2015, 2018). The SNP density of sub-genome R was only 1/6 that of the A or B sub-genomes, implying that fewer rye cultivars were involved in triticale formation. Correspondingly, 55 MTAs were in sub-genome R, while 105 and 93 were located in the sub-genomes A and B. Of 253 marker-trait associations (MTAs), only 21 reliable MTAs could be identified over two environments and three reliable MTAs over three environments. These should be derived from the complex climate in Qinghai-Tibet Plateau, low-accuracy genotype calling at some loci (Browning and Yu 2009) or small population size (Finno et al. 2014). The release of common wheat and rye genome sequences helped us to annotate the MTAs and compare our results with previous research precisely. We predicted 205 genes associated with 94 significant markers based on the wheat and rye genomes. Some annotated functional genes were relative to the related traits in previous reports (Table 1). Based on the accurate physical position, the identified loci could be differentiated clearly from

previous researches. By GBS approach, three reliable MTAs 4Brs411827353, 6Ars3762529, NewChr5rs148736541 controlling PH were identified on chromosomes 4B, 6A, and 5R, respectively. While they were not the same loci on 4B, 6A, and 5R located by previous researches (Alheit et al. 2014; Wilhelm et al. 2012; Würschum et al. 2014; Wajdzik et al. 2019) (Table S5). The identified MTAs may serve to validate and fine map the QTL in bi-parental mapping populations from dedicated crosses.

Conclusions

In summary, we conducted GBS analysis of 199 triticale cultivars and mapped reads to the A, B, D, and R genomes for karyotype analysis. Material types, abnormal behaviors, and small segment changes of some chromosomes were identified in these accessions using the read mapping method, demonstrating that this method is convenient and fast for analyzing chromosome karyotypes. The biological functions affected by abnormal chromosome behaviors need be studied further. We performed GWAS to identify significant MTAs and candidate genes. A total of 16 putative candidate genes involved in plant growth and development were acquired from the genome sequences of wheat and rye. Future studies involving gene cloning and function identification are needed to identify the relevant candidate genes controlling these traits.

Declarations

References

Akram S, Arif MAR, Hameed A (2021) A GBS-based GWAS analysis of adaptability and yield traits in bread wheat (*Triticum aestivum* L.). J Appl Genet 62:27–41

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664

Alheit KV, Reif JC, Maurer HP, Hahn V, Weissmann EA, Miedaner T, Würschum T (2011) Detection of segregation distortion loci in triticale (× *Triticosecale* Wittmack) based on a high-density DArT marker consensus genetic linkage map. BMC Genomics 12:380

Alheit KV, Busemeyer L, Liu WX, Maurer HP, Gowda M, Hahn V, Weissmann S, Reif JC, Würschum T (2014) Multiple-line cross QTL mapping for biomass yield and plant height in triticale (× *Triticosecale* Wittmack). Theor Appl Genet 127:251–260

Ayaad M, Han ZM, Zheng K, Hu G, Abo-Yousef M, Sobeih SES, Xing YZ (2021) Bin-based genome-wide association studies reveal superior alleles for improvement of appearance quality using a 4-way MAGIC population in rice. J Adv Res 28:183–194

Bai C, Liang Y, Hawkesford MJ (2013) Identification of QTLs associated with seedling root traits and their correlation with plant height in wheat. J Exp Bot 64:1745–1753

Bates D, Machler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48

Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. Plant Cell 19:148–162

Blum A (2014) The abiotic stress response and adaptation of triticale. Cereal Res Commun 42:359–375

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 Phytol 215:1026–1038

Browning BL, Yu Z (2009) Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. Am J Hum Genet 85:847–861

Cassani E, Bertolini E, Badone FC, Landoni M, Gavina D, Sirizzotti A, Pilu R (2009) Characterization of the first dominant dwarf maize mutant carrying a single amino acid insertion in the VHYNP domain of the *dwarf8* gene. Mol Breeding 24:375–385

Cheng ZJ, Murata M (2002) Triticale: a promising addition to the worldis cereal grains. Genes Genet Syst 77:23–29

Chono M, Honda I, Zeniya H, Yoneyama K, Saisho D, Takeda K, Takatsuto S, Hoshino T, Watanabe Y (2003) A semidwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiol 133:1209–1219

Costa CT, Albuquerque ACS, Nascimento Junior A, Marcelino FC, Pereira JF (2007) Genetic diversity of Brazilian triticales evaluated with genomic wheat microsatellites. Pesqui Agropecu Bras 42:1577–1586

Dennis ES, Gerlach WL, Peacock WJ (1980) Identical polypyrimidine-polypurine satellite DNAs in wheat and barley. Heredity 44:349–366

Dockter C, Gruszka D, Braumann I, Druka A, Druka I, Franckowiak J, Gough SP, Janeczko A, Kurowska M, Lundqvist J, Lundqvist U, Marzec M, Matyszczak I, Müller AH, Oklestkova J, Schulz B, Zakhrabekova S, Hansson M (2014) Induced variations in brassinosteroid genes define barley height and sturdiness, and expand the green revolution genetic toolkit. Plant Physiol 166:1912–1927

Dou QW, Tanaka H, Nakata N, Tsujimoto H (2006) Molecular cytogenetic analyses of hexaploid lines spontaneously appearing in octoploid Triticale. Theor Appl Genet 114:41–47

Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE (2008) Establishing an adjusted *p* value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics 9:516

Fan CC, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang QF (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

FAOSTAT (2017) Statistics division of food and agriculture organization of the United Nations. Available from http://faostat3.fao.org/browse/Q/QC/E, Accessed on May 8th

Finno CJ, Aleman M, Higgins RJ, Madigan JE, Bannasch DL (2014) Risk of false positive genetic associations in complex traits with underlying population structure: a case study. Vet J 202:543–549

Fox PN, Skovmand B, Thompso BK, Braun HJ, Cormier R (1990) Yield and adaptation of hexaploid spring triticale. Euphytica 47:57–64

Hao M, Luo JT, Zhang LQ, Yuan ZW, Yang YW, Wu M, Chen WJ, Zheng YL, Zhang HG, Liu DC (2013) Production of hexaploid triticale by a synthetic hexaploid wheat-rye hybrid method. Euphytica 193:347–357

Henriques R, Magyar Z, Bögre L (2013) S6K1 and E2FB are in mutually antagonistic regulatory links controlling cell growth and proliferation in *Arabidopsis*. Plant Signal Behave 8:e24367

Huber DP, Philippe RN, Madilao LL, Sturrock RN, Bohlmann J (2005) Changes in anatomy and terpene chemistry in roots of Douglas-fir seedlings following treatment with methyl jasmonate. Tree Physiol 25:1075–1083

Kanno Y, Oikawa T, Chiba Y, Ishimaru Y, Shimizu T, Sano N, Koshiba T, Kamiya Y, Ueda M, Seo M (2016) *AtSWEET13* and *AtSWEET14* regulate gibberellin-mediated physiological processes. Nat Commun 7:13245

Khalil HB, Ehdaeivand MR, Xu Y, Laroche A, Gulick PJ (2015) Identification and characterization of rye genes not expressed in allohexaploid triticale. BMC Genom 16:1–11

Komuro S, Endo R, Shikata K, Kato A (2013) Genomic and chromosomal distribution patterns of various repeated DNA sequences in wheat revealed by a fluorescence in situ hybridization procedure. Genome 56:131–137

Kuleung C, Baenziger PS, Kachman SD, Dweikat I (2006) Evaluating the genetic diversity of triticale with wheat and rye SSR markers. Crop Sci 46:1692–1700

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

Li P, Chen J, Wu P, Zhang J, Chu C, See D, Brown-Guedira G, Zemetra R, Souza E (2011) Quantitative trait loci analysis for the effect of *Rht-B1* dwarfing gene on coleoptile length and seedling root length and number of Bread Wheat. Crop Sci 51:2561–2568

Li MX, Yeung JMY, Cherny SS, Shan PC (2012) Evaluating the effective numbers of independent tests and significant *p* value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet 131:747–756

Li C, Wang P, Menzies NW, Lombi E, Kopittke PM (2018) Effects of methyl jasmonate on plant growth and leaf properties. J Plant Nutr Soil Sci 181:409–418

Li YF, Zeng XQ, Zhuang H, Chen H, Zhang T, Zhang J, Zheng H, Tang J, Wang HL, Ren SX, Ling YH, He GH (2019) Characterization and fine mapping of nonstop glumes 2 (nsg2) mutant in rice (*Oryza sativa* L.). Plant Biotechnol 36:125–134

Li C, Tao RF, Li Y, Duan MH, Xu JH (2020) Transcriptome analysis of the thermosensitive genic male-sterile line provides new insights into fertility alteration in rice (*Oryza sativa*). Genomics 112:2119–2129

Liu WX, Gowda M, Reif JC, Hahn V, Ruckelshausen A, Weissmann EA, Maurer HP, Würschum T (2014) Genetic dynamics underlying phenotypic development of biomass yield in triticale. BMC Genomics 15:458

Liu WX, Leiser WL, Reif JC, Tucker MR, Losert D, Weissmann S, Hahn V, Maurer HP, Würschum T (2016) Multiple-line cross QTL

mapping for grain yield and thousand kernel weight in triticale. Plant Breed 135:567–573

Lou GM, Chen PL, Zhou H, Xiong JW, Wan SS, Zheng YY, Alam M, Liu RJ, Zhou Y, Yang HY, Tian YH, Bai JL, Rao WT, Tan X, Gao HZ, Li YH, Gao GJ, Zhang QD, Li XH, Liu CG, He YQ (2021) FLOURY ENDOSPERM19 encoding a class I glutamine amidotransferase affects grain quality in rice. Mol Breed 41:1–15

Lukaszewski AJ, Gustafson JP (1987) Cytogenetics of Triticale. Plant Breed Rev 5:41–93

Ma XF, Gustafson JP (2008) Allopolyploidization-accommodated genomic sequence changes in triticale. Ann Bot-London 101:825–832

McGoverin CM, Snyders F, Muller N, Botes W, Fox G, Manley M (2011) A review of triticale uses and the effect of growth environment on grain quality. J Sci Food Agr 91:1155–1165

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. ASA-CSSA, Madison, pp 99–118

Mergoum M, Gómez-Macpherson H (2004) Triticale improvement and production FAO Plant Production and Protection Papers Food and Agriculture Organization of the United Nations. Food & Agriculture Org, Rome, p 179

Niedziela A, Orłowska R, Machczyńska J, Bednarek PT (2016) The genetic diversity of triticale genotypes involved in Polish breeding programs. Springerplus 5:1–7

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Rabanus-Wallace MT, Hackauf B, Mascher M, Lux T, Wicker T, Gundlach H, Baez M, Houben A, Mayer KFX, Guo LL, Poland J, Pozniak CJ, Walkowiak S, Melonek J, Praz CR, Schreiber M, Budak H, Heuberger M, Steuernagel B, Wulff B, Börner A, Byrns B, Čížková J, Fowler DB, Fritz A, Himmelbach A, Kaithakottil G, Keilwagen J, Keller B, Konkin D, Larsen J, Li Q, Myśków B, Padmarasu S, Rawat N, Sesiz U, Biyiklioglu-Kaya S, Sharpe A, Šimková H, Small I, Swarbreck D, Toegelová H, Tsvetkova N, Voylokov AV, Vrána J, Bauer E, Bolibok-Bragoszewska H, Doležel J, Hall A, Jia JZ, Korzun V, Laroche A, Ma XF, Ordon F, Özkan H, Rakoczy-Trojanowska M, Scholz U, Schulman AH, Siekmann D, Stojałowski S, Tiwari VK, Spannagl M, Stein N (2021) Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. Nat Genet 53:564–573

Rahimi Y, Bihamta MR, Taleei A, Alipour H, Ingvarsson PK (2019) Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. BMC Plant Biol 19:1–19

Sakiroglu M, Brummer EC (2017) Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. Theor Appl Genet 130:261–268

Singh M, Ceccarelli S, Hamblin S (1993) Estimation of heritability from varietal trials data. Theor Appl Genet 86:437–441

Sun TP (2008) Gibberellin metabolism, perception and signaling pathways in *Arabidopsis*. Arabidopsis Book 6:e0103

Tams SH, Bauer E, Oettler G, Melchinger AE (2004) Genetic diversity in European winter triticale determined with SSR markers and coancestry coefficient. Theor Appl Genet 108:1385–1391

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evolut 28:2731–2739

Tang Z, Yang Z, Fu S (2014) Oligonucleotides replacing the roles of repetitive sequences pAs1, pSc119.2, pTa-535, pTa71, CCS1, and pAWRC.1 for FISH analysis. J Appl Genet 55:313–318

Torkamaneh D, Jérme L, Boyle B, Hyten DL, Belzile F (2021) A bumper crop of SNPs in soybean through high-density genotyping-by-sequencing (HD-GBS). Plant Biotechnol J 19:860–862

Tyrka M, Bednarek PT, Kilian A, Wędzony M, Hura T, Bauer E (2011) Genetic map of triticale compiling DArT, SSR, and AFLP markers. Genome 401:391–401

Tyrka M, Tyrka D, Wędzony M (2015) Genetic map of triticale integrating microsatellite DArT and SNP markers. PLoS ONE 10:e0145714

Tyrka M, Oleszczuk S, Rabiza-Swider J, Wos H, Wedzony M, Zimny J, Ponitka A, Ślusarkiewicz-Jarzina A, Metzger RJ, Baenziger PS (2018) Populations of doubled haploids for genetic mapping in hexaploid winter triticale. Mol Breeding 38:46

Wajdzik K, Gołębiowska G, Dyda M, Gawrońska K, Rapacz M, Wędzony M (2019) The QTL Mapping of the important breeding traits in winter triticale (× *Triticosecale* Wittmack). Cereal Res Commun 47:395–408

Wang SK, Wu K, Yuan QB, Liu XY, Liu ZB, Lin XY, Zeng RZ, Zhu HT, Dong GJ, Qian Q, Zhang GQ, Fu XD (2012) Control of grain size, shape and quality by OsSPL16 in rice. Nat Genet 44:950

Wilhelm EP, Mackay IJ, Saville RJ, Korolev AV, Balfourier F, Greenland AJ, Boulton MI, Powell W (2012) Haplotype dictionary for the *Rht-1* loci in wheat. Theor Appl Genet 126:1733–1747

Wu YF, Lee SK, Yoo Y, Wei JH, Kwon SY, Lee SW, Jeon JS, An G (2018) Rice Transcription Factor *OsDOF11* modulates sugar transport by promoting expression of sucrose transporter and *SWEET* genes. Mol Plant 11:833–845

Würschum T, Liu WX, Busemeyer L, Tucker MR, Reif JC, Weissmann EA, Hahn V, Ruckelshausen A, Maurer HP (2014) Mapping dynamic QTL for plant height in triticale. BMC Genet 15:59

Yan ZH, Wan YF, Liu KF, Zheng YL, Wang DW (2002) Identification of a novel HMW-glutenin subunit and comparison of its amino acid sequence with those of homologous subunits. Sci Bull 47:222–226

Yang WN, Guo ZL, Huang CL, Duan LF, Chen GX, Jiang N, Fang W, Feng H, Xie WB, Lian XM, Wang GW, Luo QM, Zhang QF, Liu Q, Xiong LZ (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat Commun 5:5087

Yang JH, Ding CQ, Xu BC, Chen CT, Narsai R, Whelan J, Hu ZY, Zhang MF (2015) A Casparian strip domain-like gene, *CASPL*, negatively alters growth and cold tolerance. Sci Rep 5:14299

Yang MM, Dong J, Zhao WC, Gao X (2016) Characterization of proteins involved in early stage of wheat grain development by iTRAQ. J Proteomics 136:157–166

Zhang Y, Xiong Y, Liu RY, Xue HW, Yang ZB (2019) The Rho-family GTPase OsRac1 controls rice grain size and yield by regulating cell division. P Natl Acad Sci USA 116:16121–16126

Zhao LB, Ning SZ, Yi YJ, Zhang LQ, Yuan ZW, Wang JR, Zheng YL, Hao M, Liu DC (2018) Fluorescence in situ hybridization karyotyping reveals the presence of two distinct genomes in the taxon *Aegilops tauschii*. BMC Genomics 19:1–9

Zhu F (2018) Triticale: Nutritional composition and food uses. Food Chem 241:468–479