Genomic Regions Controlling Yield-related Traits in Spring Wheat: A Mini Review and a Case Study for Rainfed Environments in Australia and China

Hui Liu
  The University of Western Australia

Daniel Mullan
  InterGrain Pty Ltd

Shancen Zhao
  Beijing Genomics Institute: BGI Group

Yong Zhang
  Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Jun Ye
  Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences

Yong Wang
  Gansu Academy of Agricultural Sciences

Aimin Zhang
  Institute of Genetics and Developmental Biology CAS: Institute of Genetics and Developmental Biology Chinese Academy of Sciences

Xiaoqing Zhao
  Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences

Guannan Liu
  The University of Western Australia

Chi Zhang
  Beijing Genomics Institute: BGI Group

Kenneth Chan
  AGRF: Australian Genome Research Facility Ltd

Zhanyuan Lu
  Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences

Guijun Yan (✉ guijun.yan@uwa.edu.au)
  The University of Western Australia   https://orcid.org/0000-0001-9628-1211

Research Article

Abstract

Breeding high-yield wheat varieties performing well in target environment is economically important. This study conducted a mini review of genome-wide association study (GWAS) outcomes on wheat yield-related traits reported in recent years, and performed GWAS in six individual environments to identify major alleles and their candidate genes responsible for wheat yield-related traits in Australia and North China where rainfed farming system is adopted. A panel of 228 spring wheat varieties were used. A double digest restriction-site associated DNA (ddRAD) genotyping-by-sequencing (GBS) protocol was performed to generate single nucleotide polymorphism (SNP) marker data. A total of 223 significant marker-trait association (MTAs) for yield traits, and 46 candidate genes for the major or consistent MTAs were identified. A phenomenon seldom reported in previous studies was that MTA clustered chromosome segments and gene clusters responsible for the trait were found across the genome, which suggested that marker-assisted selection (MAS) or transgenic method targeting a single gene might not be as effective as MAS targeting a much larger genomic region (GR) where all the genes or gene clusters underlying the GR play important roles.

Key Messages

Gene clusters and clustered marker-trait associations responsible for wheat yield were found from different genome-wide associationstudies and meta-QTL analyses.

Introduction

As a major food crop, wheat has long been a target by researchers for production improvement, especially in recent years when adverse weather has increased resulting from climate change. Like most crops, wheat production is highly dependent on individual environment, and therefore needs to be analysed individually in specific environments. In many areas of the world, wheat is grown under rainfed conditions. Climate change has resulted in increased frequency and intensity of extreme conditions such as drought and heat in these areas, which can significantly reduce wheat yields. Therefore, to meet predicted global demand, wheat yield under rainfed condition has attracted a wide research interest.

In recent years, genome-wide association study (GWAS) is widely used to identify genomic regions (GRs) responsible for yield-related traits, apart from quantitative trait locus (QTL) linkage mapping. Compared to bi-parental QTL mapping, GWAS can capture more GRs responsible for the traits due to diversity of the association panels (Alqudah et al. 2020; Sonah et al. 2015). The diverse and unstructured germplasm used in GWAS have accumulated a larger number of recombination events, therefore, accompanied with high resolution markers they can increase mapping accuracy and shorten confidence intervals of the responsible GRs (Wu et al. 2020). GWAS may also allow candidate genes to be identified if a large association mapping population is used (Pang et al. 2020). With the development of next generation sequencing (NGS) technology and the publication of wheat reference genome (Appels et al. 2018), GWAS has identified many GRs and candidate genes which had not been detected in QTL linkage mapping (Juliana et al. 2019). However, unlike QTL linkage mapping where several meta-QTL (MQTL) analyses were conducted to integrate the identified GRs for yield-related traits in wheat (Acuna-Galindo et al. 2015; Liu et al. 2020; Zhang et al. 2010), there is no review on yield-related GRs identified by GWAS although numerous such studies have been published (Table S1).

Among the high-resolution genotyping methods, genotyping by sequencing (GBS) and SNP arrays are the most commonly adopted, where GBS has more power to detect rare alleles in diverse germplasm collections and SNP array has the advantage of lower cost per data-point (Darrier et al. 2019). Along with the reduced cost of NGS, GBS has been increasingly adopted in genetic studies, in which double digest restriction-site associated DNA sequencing (ddRADseq) is one of the promising options because it can achieve relatively higher genome coverage — the marker number can be customized depending on the choice of enzymes and range of fragment sizes selected (Cumer et al. 2021). If multiple traits are simultaneously analysed, the GWAS cost can be significantly reduced.

The major wheat production areas in Australia have either Mediterranean climate zone such as Western Australia (WA) and South Australia (SA), or temperate climate zone such as New South Wales (NSW) and Victoria (VIC), where wheat is mostly grown in winter season and the crops may encounter drought and heat when approaching to harvest. Spring wheat with no vernalization requirement is the main wheat type used. In North China such as Inner Mongolia (IM) and Gansu, spring type wheat is also widely cultivated. North China generally has a temperate monsoon or temperate continent climate, and spring wheat is harvested in summer when the weather can be hot and dry. Australia and North China have very different climates, but rainfed farming system is adopted for spring wheat in both areas, which makes it interesting to investigate if there is any consistency of GRs for yield-related traits between the two places.

The objectives of this study are: 1) summarizing the GRs for wheat yield-related traits reported in previous GWAS and MQTL analyses; 2) conducting GWAS on wheat yield-related traits under rainfed condition in different environments; 3) comparing the GRs among the different environments and with the reported GRs; and 4) identifying key GRs and their candidate genes responsible for wheat yield.

Materials And Methods

Review of reported GRs for wheat yield-related traits identified by GWAS

Twenty-three GWAS publications on wheat yield-related traits were collected (Table S1), and their identified GRs were summarized. Generally, those with a significance threshold of $-\log_{10} (p) > 4$ were selected, and those showed consistency within the study or with previously reported GRs were also included. For GRs in some studies where the marker physical positions were not indicated due to the lack of reference genome information at the time of publication, their available SNP sequences (either from known SNPs such as wheat 90K Illumina iSelect array or from the marker sequences given in the paper) were searched for sequence similarity using BLAST against the Chinese Spring (CS) reference genome, International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.0, to locate their physical positions (https://wheat-urgi.versailles.inra.fr/) (Appels et al. 2018). A MQTL study of GRs for wheat yield traits was also incorporated, where the physical positions of the major GRs were indicated (Liu et al. 2020). The distribution of the summarized GRs on wheat chromosomes were visualized using R package RIdeogram (Hao et al. 2020).

Plant materials, phenotyping and statistical analyses of phenotypic data

A total of 228 diverse spring wheat (*Triticum aestivum* L.) varieties were used as the association panel and were field-trialled under six different environments in Australia and China (Table S2). Most of these varieties have long been cultivated in either of the countries. Among them, 150 lines were trialled at The University of Western Australia (UWA) Shenton Park Field Station (31.9588° S, 115.8053°E) in 2019, and in different Australian fields at Brocklesby, NSW (35.8738°S, 146.6817° E), Cunderdin, WA (31.4926°S, 117.4928°E), Lock, SA (33.5973°S, 135.7110°E) and Kalkee, VIC (36.5640°S, 142.1874°E) in 2020; and 133 lines were trialled at the Inner Mongolia (IM) Academy of Agricultural & Animal Husbandry Sciences (IMAAAS) (43.3782° N, 115.0595° E) in 2019. There were 55 overlapped varieties between the Australian trials and the IM trial. Seeds were obtained from Australian Winter Cereals Collection and wheat breeding companies/institutions including InterGrain Pty Ltd, Australian Grain Technologies, LongReach Plant Breeders, Edstar Genetics Pty Ltd, Chinese Academy of Agriculture Sciences, IMAAAS, and Gansu Academy of Agricultural Sciences. Field was ploughed two weeks prior to seed sowing and fertilizers were added with standard rates. All plots were under rainfed conditions.

Not all the traits were recorded in all the environments. In UWA Shenton Park trial, three blocks, each with 150 plots for 150 varieties (Table S2), were used as three replicates for the individual varieties which were planted in generalized randomized block design. Each plot had a size of 1.0 m × 0.6 m with three rows. Sixty seeds were sown in each plot, with 20 seeds (spaced 5 cm apart) in each 1 m row. Distance between rows was 0.2 m, and distance between plots was 0.4 m. Buffer zones were two rows surrounding the whole field. Seeds were sown in mid-May in 2019. Trials were phenotyped for plant height (PH), above ground biomass per plant (BM), spike number per plant (SN), grain number per plant (GN), thousand

grain weight (TGW) and yield (YLD) per square meter. PH was recorded as the average of three values for each plot measured in centimeters from the soil surface to the tip of the spike excluding awns. BM was measured as above ground crop cuts of each line, which were dried in a 60°C oven before being weighed to assess the dry biomass. SN, GN, TGW and YLD were counted or measured at maturity when whole plots were harvested.

In 2020 Australian trials, seeds of 150 varieties (Table S2) were sown at different locations in WA, SA, NSW and VIC in mid-May, 2020. A randomized complete block design was used for all locations. At Brocklesby (NSW), each variety was grown in a 1.14 m × 4 m plot with two replications; at Cunderdin (WA), each variety was grown in a 1.52 m × 4 m plot with two replications; at Lock (SA), each variety was grown in a 1.65 m × 4 m plot with two replications; and at Kalkee (VIC), each variety was grown in a 1 m × 4 m plot with two replications. YLD was measured at maturity for all the trials.

In 2019 IM trials, seeds of 133 varieties (Table S2) were sown in late-March at IMAAAF in 2019. A randomized complete block design was used and every wheat line was planted in a 1.2 m × 0.2 m plot with three replicates. Fifty seeds were sown in each plot. Five buffer lines were used for the whole field. Phenotyping were done for PH, fresh BM, SN, GN, TGW and YLD per plant. Except fresh BM which were measured as fresh weight of the above ground crop cuts of each plant, other traits were measured using the similar methods as the UWA trial.

Correlations of the traits in individual environments were obtained using the Pearson's product–moment estimator. Frequency distributions and correlation plots of the traits were generated using R package PerformanceAnalytics (https://github.com/braverock/PerformanceAnalytics).

Genotyping by sequencing (GBS) and data filtering

Genomic DNAs of the panel plants were extracted from three-leaf stage seedling leaves using a modified CTAB method as described previously (Wang et al. 2019). The yield, purity and integrity of the DNAs were checked using NanoDrop 2000 (Thermo Fisher Scientific Inc., Australia) and agarose gel electrophoresis. The Illumina platform was used for genotyping at Beijing Genomics Institute (BGI), and a double digest restriction-site associated DNA (ddRAD) GBS approach was adopted to obtain genome-wide single nucleotide polymorphism (SNP) and insertion-deletion (InDel) markers for all the lines. The protocols of DNA double digestion, library construction and raw sequencing data filtering were the same as described in the previous study (Liu et al. 2020). BWA (Li and Durbin 2009) was used to map the clean reads to the reference genome IWGSC RefSeq v1.0. Marker polymorphisms were called using Samtools (Li et al. 2009), Picard-tools (http://broadinstitute.github.io/picard/), Reseqtools (He et al. 2013), and GATK unifiedGenotyper (DePristo et al. 2011).

The variant call format (vcf) data files of SNP markers generated by BGI were further filtered using VCFtools (http://vcftools.sourceforge.net) for downstream analyses with the following criteria: 1) only keeping variants that were successfully genotyped in 80% of individuals, with a minor allele frequency (MAF) > 5% and a minimum quality score of 30; 2) removing varieties with more than 30% missing data to get rid of individuals that did not sequence well; and 3) after missing data were imputed with Beagle v5.1 (Browning et al. 2018), the markers were further pruned using plink v1.9 (Purcell et al. 2007) with a cut-off linkage disequilibrium (LD) value $r^2 > 0.8$.

Population structure analyses, principal components analyses (PCA), kinship analyses and phylogenetic tree constructions

Population structure was analysed with Admixture 1.3.0 (Alexander and Lange 2011). The number of presumed subpopulations (K) was set to 2 to 15 and the cross validation (CV) errors for the different admixture models were calculated. The K value with the smallest CV error were considered the best fitting model. R package LEA (Frichot and François 2015) was used to generate structure plots for the varieties trialled in Australia and IM respectively.

The first three PCs and a kinship matrix were generated using TASSEL v5 (Bradbury et al. 2007). Neighbor-joining trees for the phylogenetic relationships of the varieties in individual environments (including Australian and IM environment) were drawn using Mega X software (Kumar et al. 2018).

Genome-wide association mapping

Genome-wide association mapping between SNP markers and phenotypic data in individual environments was conducted with TASSEL v5 using a mixed linear model (MLM) that accounts for both kinship coefficients and population structure. Marker-trait associations (MTAs) were claimed as significant with a p value threshold of − log10 > 4 for traits of PH, BM, TGW, SN and GN, and a p value threshold of − log10 > 3 for YLD. Manhattan and QQ plots were drawn using R package CMplot (Yin et al. 2021).

Candidate gene identification for major MTAs

The trait-associated SNP markers, especially those showed consistency under different environments in this study, were compared with previously reported markers/genes or MQTL near the significant MTAs, and were searched in the reference genome RefSeq v1.0 to identify their overlapped or closely located high-confidence genes as associated candidate genes. Those genes with functions previously reported as important for yield (Liu et al. 2020) were especially considered as key candidates.

Comparison of MTAs for yield to previously reported MTAs for maturity

As days to maturity (maturity in short) can significantly affect yield traits under different environments, the major (-log$_{10}$ (p) > 4) MTAs for maturity identified in a previous study (Juliana et al. 2019) were summarized and visualized using RIdeogram. These reported maturity MTAs were considered representative because they were identified across different seasons under multiple environments including irrigated, drought stressed, early heat stressed and late heat stressed conditions. The GRs for yield traits reported previously and in this study were compared to the maturity MTAs to explore the possible influence of maturity to yield traits.

Results

GRs responsible for wheat yield previously reported by GWAS and MQTL analyses

The major (-log$_{10}$ (p) > 4) or consistent GRs for wheat yield summarized from the 23 GWAS reports (Table S1) and a recent MQTL analysis (Liu et al. 2020) were shown in Fig. 1, together with the yield GRs identified in this study. A total of 506 GRs from GWAS and 86 GRs from MQTL analysis were compared. For the distribution of the compared GRs identified by GWAS, chromosome 2B had the most number (60) of markers, followed by chromosome 5B (59), whereas chromosome 1D and 5D had the least number (11) of markers, followed by chromosome 4D (12). It was evident from Fig. 1 that the majority of the markers were densely located towards the two telomere ends of the chromosomes. On some chromosomes, however, these marker clusters were also obviously observed on the other parts of the chromosomes apart from the two ends, for example, segment 388778949−405287003 bp on chromosome 2B, segment 572254295−669283233 bp on chromosome 3B, segments 464478741−497453460 bp and 555768596−618950471 bp on chromosome 5A, segment 387952822− 489300000 bp on chromosome 5B, segment 156270035−239330009 bp on chromosome 6B, and segment 81662899− 139383754 on chromosome 6D. Among them, the segments on chromosome 3B, 5A, 5B, 6B and 6D each harbours more than ten MTAs within a 60−100 Mbp physical interval (Table S1). Sixty-eight GWAS alleles/GRs were found overlapping with 27 reported MQTL, and some of those falling in the same MQTL region were from different studies (Table 1).

Table 1
Genomic regions (GRs) of yield-related traits in reported genome wide association studies (GWAS) overlapped or closely located (distance < 5Mbp) to meta-QTL (MQTL)

GWAS allele/GR (Chr:Pos)	MQTL name (interval)*	Distance to MQTL (bp)	GWAS reference
1B:539.6-542.6M	MQTL1B.1 (543082472–543517652)	0.48M	(Li et al. 2019a)
1B:561706780–561706864	MQTL1B.2 (561507423–565335030)	Within MQTL interval	(Muhu-Din Ahmed et al. 2020)
1B:645269062	MQTL1B.8 (638015656–641199314)	4.07M	(Li et al. 2019b)
1D:33989918	MQTL1D.1 (27540526–27576726)	1.13M	(Pang et al. 2020)
2A:30.3-31.9M	MQTL2A.1 (31953248–32864757)	1.65M	(Li et al. 2019a)
2A:87856661	MQTL2A.2 (79751994–84951037)	2.91M	(Li et al. 2019b)
2A:770198409	MQTL2A.3 (763873451–771073270)	Within MQTL interval	(Li et al. 2019b)
2B:44983849	MQTL2B.1 (44424318–52670041)	Within MQTL interval	(Ward et al. 2019)
2B:71692883–71692929	MQTL2B.2 (65370433–76489059)	Within MQTL interval	(Sehgal et al. 2017)
2B:388778949; 2B:391365352; 2B:399704597; 2B:400995689; & 2B:405287003	MQTL2B.3 (263220312–453816238)	Within MQTL interval	(Pang et al. 2020); (Li et al. 2019b); (Juliana et al. 2019)
2B:689481217–689481285; 2B:707958872; 2B:726053574; 2B:744853702; 2B:747226742; 2B:750833541; 2B:775174987; & 2B:775.83-777.51M	MQTL2B.5 (683043641–779229586)	Within MQTL interval	(Sehgal et al. 2017); (Li et al. 2019b); (Li et al. 2019c); (Liu et al. 2018)
2D:52106277	MQTL2D.2 (29715676–50941419)	1.16M	(Li et al. 2019b)
2D:83138642	MQTL2D.4 (75393708–79941414)	3.20M	(Pang et al. 2020)
3A:466777084; & 3A:479477047–479477147	MQTL3A.3 (302928129–480147815)	Within MQTL interval	(Juliana et al. 2019); (Muhu-Din Ahmed et al. 2020)

*MQTL names are from a previous meta-QTL study (Liu et al. 2020).

Chr: Chromosome; Pos: physical position (bp); M: million.

GWAS allele/GR (Chr:Pos)	MQTL name (interval)*	Distance to MQTL (bp)	GWAS reference
3B:7031744−7031759	MQTL3B.1 (5673703−8814393)	Within MQTL interval	(Sehgal et al. 2020)
3B:20.5-22.0M	MQTL3B.2 (21343759−23600280)	Partly overlap with MQTL	(Li et al. 2019b)
3B:151991918; & 3B:154399597−155399597	MQTL3B.4 (140851970−257841306)	Within MQTL interval	(Pang et al. 2020); (Li et al. 2019c)
3B: 326243214	MQTL3B.5 (242168620−414186365)	Within MQTL interval	(Pang et al. 2020)
3B:781549015−782283624	MQTL3B.6 (779535677−783472564)	Partly overlap with MQTL	(Juliana et al. 2019)
3D: 85853075−86853075	MQTL3D.1 (51042786−86353210)	Partly overlap with MQTL	(Li et al. 2019c)
4A:719934001−719934069; & 4A:722069775	MQTL4A.1 (717297782−722909098)	Within MQTL interval	(Sehgal et al. 2017); (Juliana et al. 2019)
4A: 665504094	MQTL4A.2 (660988814−666149150)	Within MQTL interval	(Pang et al. 2020)
4A:681361201; 4A:681651206; & 4A:682741218	MQTL4A.3 (673446685−684269347)	Within MQTL interval	(Li et al. 2019b)
4A:632152511; 4A:642368219; 4A:643146184−646088197; 4A:647346724; 4A:648132716; 4A:665504094; 4A:667344925−667344993; 4A:681361201; 4A:681651206; & 4A:682741218	MQTL4A.4 (629917641−705760459)	Within MQTL interval	(Pang et al. 2020); (Li et al. 2020); (Juliana et al. 2019); (Li et al. 2019b); (Gahlaut et al. 2019)
4B:28740074	MQTL4B.2 (27396703−28954608)	Within MQTL interval	(Pang et al. 2020)
4B:37529691−37529759	MQTL4B.3 (33613374−35515028)	2.01M	(Gahlaut et al. 2019)
4B:159.2-163.0M; & 4B:320327920−320327921	MQTL4B.4 (132334183−409740551)	Within MQTL interval	(Li et al. 2019a); (Li et al. 2019b)
4B:535080031−535080131	MQTL4B.7 (519257662−531053980)	4.03M	(Ain et al. 2015)

*MQTL names are from a previous meta-QTL study (Liu et al. 2020).

Chr: Chromosome; Pos: physical position (bp); M: million.

GWAS allele/GR (Chr:Pos)	MQTL name (interval)*	Distance to MQTL (bp)	GWAS reference
5A:464478741−464478841; & 5A:466985296−466985414	MQTL5A.3 (461519115−470033346)	Within MQTL interval	(Qaseem et al. 2019)
5A:595083572	MQTL5A.2 (592280059−594962156)	2.80M	(Li et al. 2019b)
5A:616879504−616879604	MQTL5A.1 (617352363−625177133)	4.73M	(Sukumaran et al. 2015)
5B:38166696−38166796	MQTL5B.1 (37278118−78551199)	Within MQTL interval	(Qaseem et al. 2019)
6A:562931571−562931671	MQTL6A.2 (562931571−563378468)	Within MQTL interval	(Sun et al. 2017)
6B:50987729−50987829; 6B:50990383−50990483; 6B:59294578; & 6B:60597908	MQTL6B.2 (40456055−77011442)	Within MQTL interval	(Sun et al. 2017); (Muhu-Din Ahmed et al. 2020); (Li et al. 2019b)
6D:16854225	MQTL6D.4 (17257264−24003436)	4.03M	(Li et al. 2019b)
7A:586312360	MQTL7A.3 (581349401−611333731)	Within MQTL interval	(Juliana et al. 2019)
7B:444463299; 7B:451655671; 7B:568676196; 7B:575475604−575475775; 7B:604038725−604038793; 7B:639791862; 7B:642758455−642758555; 7B:649370094; 7B:660237170; 7B:699838532−699838632; 7B:701871809−701872008; 7B:702233717−702233817; & 7B:703266182−703266250	MQTL7B.4 (444475000−678635377); & MQTL7B.2 (648658065−713633034)	Within MQTL intervals	(Ward et al. 2019); (Muhu-Din Ahmed et al. 2020); (Li et al. 2020); (Sehgal et al. 2017); (Qaseem et al. 2019); (Pang et al. 2020); (Li et al. 2019b)
7D:63.0-69.7M	MQTL7D.3 (59091469−63444566)	Within MQTL interval	(Li et al. 2019a)
*MQTL names are from a previous meta-QTL study (Liu et al. 2020).			
Chr: Chromosome; Pos: physical position (bp); M: million.			

Genotype data sorting and phenotypic variations

A total of 283,858 SNP markers were generated from ddRAD sequencing of the 228 investigated varieties. For GWAS, a pruned subset of 36,934 and 43,616 markers with missing data less than 20% and LD value $r^2 < 0.8$ were imputed and used for the Australian trial population and the IM trial population, respectively (Suppl data 1 and 2). After filtering out the varieties with more than 30% missing data, a total of 212 varieties were used for final analysis. Among them, 134 lines were trialled in Australia (2019 Shenton Park; and 2020 Brocklesby, Cunderdin, Lock and Kalkee); and 133 lines were trialled in the field at IMAAAS, China.

The yield-related traits showed abundant phenotypic variations in individual environments (Table 2 and S3). The Pearson's correlation coefficients for the six yield-related traits ranged from 0.02 (PH and GN) to 0.73 (BM and YLD) in the UWA

Shenton Park trial, and from 0.03 (PH and GN) to 0.86 (GN and YLD) in the IM trial (Fig. S1). BM and GN showed significant ($p < 0.05$) or highly significant ($p < 0.001$) positive correlations with YLD in both UWA and IM trials.

Table 2
Phenotypic variations of the yield-related traits in the trials of six environments

Australian trial (2019 & 2020)*					Chinese trial (2019)				
Trait_Environment	Max	Min	Mean	SD	Trait_Environment	Max	Min	Mean	SD
YLD_Brocklesby (t/ha)	6.73	0.22	5,22	1.10	YLD_IM (g/plant)	11.75	2.50	6.86	1.78
YLD_Cunderdin (t/ha)	1.12	0.01	0.61	0.26	PH_IM (cm)	106.67	46.42	76.26	11.58
YLD_Lock (t/ha)	2.45	0.29	1.62	0.43	BM_IM (fresh) (g)	42.33	11.90	21.97	5.20
YLD_Kalkee (t/ha)	7.39	0.37	5.52	1.18	SN_IM	10.17	2.67	5.38	1.43
YLD_UWA (g/m^2)	975.38	240.19	678.39	118.58	GN_IM	307.17	73,83	176.09	37.50
PH_UWA (cm)	125.50	55.67	76.68	11.27	TGW_IM (g)	49.98	26.75	38,68	5.61
BM_UWA (dry) (g)	18.58	5,58	10.54	1,99					
SN_UWA	6.58	1,75	3.65	1.01					
GN_UWA	158.65	46.67	98.81	25.11					
TGW_UWA (g)	96.03	22.77	46.84	8.80					
*The Brocklesby, Cunderdin, Lock and Kalkee trials were conducted in 2020, and the UWA trials were conducted in 2019.									
SD: standard deviation.									

Population structure

The number of subpopulations k = 7 was determined by Admixture software for the 134 population lines in the Australian trials and the 133 lines in the Chinese trial (Fig. 2). PCA showed that the first three PCs explained 9.57%, 6.88% and 5.46% of the genetic variance in the 134 Australian trial population, and explained 11.1%, 6.67% and 5.30% of the genetic variance in the 133 IM trial population, respectively. The cluster analyses using the first two PCs showed that the genotypes were classified into three groups in both populations (Fig. 2). In the Australian trial population, varieties Arrow, Bremer, Chief, Corack, Havoc, Impress, Maze, Ninja, Scepter, Wallup, Wyalkatchem and Zen were clustered into a separate group from the majority of the Australian varieties, while in the IM trial population, the varieties collected from Heilongjiang Province were clustered into a separate group from the majority of the Chinese varieties. Phylogenetic analyses using neighbour-joining trees (Fig. S2) supported the results from the structure and PCA analyses.

MTAs for yield-related traits

A total of 223 MTAs for YLD were identified with a significant threshold of $-\log_{10}$ (p) > 3 in different environments (Fig. 3; Table S4), and 39 of them showing major or consistent effects were shown in Table 3. Those identified in one environment and closely located, with a physical distance generally less than 10 Mbp between each other, were considered as one locus (Table S4). Chromosome distribution of these loci, together with other identified MTAs, were shown in Fig. 1 (as warm coloured dots). Among the major or consistent MTAs, twenty-three had a significant threshold of $-\log_{10}$ (p) > 4, with twenty-two identified in Australian environment and one in IM environment. The MTAs showing consistent effects on multiple traits included SNPs 1B:212877972, 2B:122702483, 3B:115783172 and 3D:3034125 which were significant for both YLD and GN;

SNP 2A:1269018 which was significant for both YLD and TGW; and SNPs 3B:208548740 and 3B:217005195 which were significant for three traits of YLD, GN and BM. MTAs that were consistent across environments/years were also identified. The consistency was mostly found in 2020 Australian trials: SNPs 1A:465076391, 1B:215325944 and 1B:215432253 were detected at both Brocklesby and Lock; SNPs 1A:465467894, 2B:287801761, 3D:334295007, 4B:209651192, 5A:248358792 and 7B:174805041 were detected at both Brocklesby and Kalkee; SNP 2D:34400629 was detected at both Brocklesby and Cunderdin; SNP 6B:17675469 was detected at both Cunderdin and Kalkee; and SNP 7B:43893059 was detected at both Cunderdin and Lock. One SNP 1B:406281153 was detected across two years at both 2019 Shenton Park and 2020 Cunderdin trials. The phenotypic variation explained (R^2) by the major or consistent MTAs ranged from 1.97% by SNP 1A:465076391 to 19.23% by SNP 5A:33961012.

Table 3
Major or consistent marker-trait associations (MTAs) responsible for yield identified by GWAS in Australian and Chinese trials

SNP (chromosome: physical position)	Significance threshold* (-log $_{10}$ (p))	Additive effect	R^2 (%)**	Environment/year	Comparison with previously reported GRs***
1A:465467894	4	-0.79	16.25	Brocklesby & Kalkee/2020	
2B:144139063	4	55.99	14.96	Shenton Park/2019	
2B:287801761	4	-	12.33	Brocklesby & Kalkee/2020	Within MQTL2B.3
3D:145929682	4	1.12	16.55	Kalkee/2020	
3D:334295007	4	-	14.23	Brocklesby & Kalkee/2020	
4B:209651192	4	-	7.42	Brocklesby & Kalkee/2020	Within MQTL4B.6
5A:29297520	4	-	17.29	Kalkee/2020	7.88 Mbp away from an MTA in (Li et al. 2020)
5A:29713379	4	-	17.29	Kalkee/2020	
5A:30654456	4	-0.84	17.38	Kalkee/2020	
5A:30691578	4	-	17.29	Kalkee/2020	
5A:32194587	4	-	17.29	Kalkee/2020	
5A:32571863	4	-	17.29	Kalkee/2020	
5A:32982772	4	0.81	18.34	Kalkee/2020	
5A:33961012	4	0.77	19.23	Kalkee/2020	
5A:237084603	4	-0.60	9.81	Brocklesby/2020	
5A:248358792	4	1.45	14.87	Brocklesby & Kalkee/2020	2.59 Mbp away from an MTA in (Li et al. 2019b)
5B:135262608	4	44.39	14.86	Shenton Park/2019	
5D:95706398	4	1.07	10.90	Brocklesby/2020	7.77 Mbp away from an MTA in (Pang et al. 2020)
7A:34077500	4	-26.20	14.04	Shenton Park/2019	
7B:224008352	4	-0.04	17.71	IM/2019	

*MTAs with -log $_{10}$ (p) > 4 were considered as major MTAs.

**If MTAs showed significance in more than one location, the maximum R^2 is shown.

***Only those reported GRs with distance of less than 10 Mbp to the SNPs identified in this study are shown. MQTL names are according to a previous meta-QTL study (Liu et al. 2020).

SNPs with underlines indicate those that were considered as one locus on their individual chromosomes, because they were identified in one environment and were closely located (with a physical distance < 10 Mbp between each other).

SNP (chromosome: physical position)	Significance threshold* (-log $_{10}$ (p))	Additive effect	R^2 (%)**	Environment/year	Comparison with previously reported GRs***
7B:155982479	4	-0.08	16.13	Kalkee/2020	4.63 Mbp away from an MTA in (Sehgal et al. 2020)
7B:176068517	4	-	12.06	Kalkee/2020	
7B:176105704	4	-0.77	15.16	Kalkee/2020	
1A:35952911	3			Shenton Park/2019	0.85 Mbp away from an MTA in (Li et al. 2019a)
1A:465076391	3	-	1.97	Brocklesby & Lock/2020	
1B:212877972 (YLD & GN)	3	0.94	12.06	IM, China/2019	
1B:215325944	3	-	12.69	Brocklesby & Lock/2020	
1B:215432253	3	-	12.69	Brocklesby & Lock/2020	
1B:406281153	3	-0.23	12.52	Cunderdin/2020 & Shenton Park/2019	
2A:1269018 (YLD & TGW)	3	0.72	11.45	IM, China / 2019	5.51 and 6.03 Mbp away from two MTAs in (Juliana et al. 2019)
2B:122702483 (YLD & GN)	3	-31.39	11.42	Shenton Park/2019	6.31 Mbp away from (Juliana et al. 2019)
2D:34400629	3	-0.16	13.17	Brocklesby & Cunderdin/2020	Within MQTL2D.1
3B:115783172 (YLD & GN)	3	-0.82	11.70	IM, China / 2019	
3B:208548740 (YLD & GN & BM)	3	-	11.93	IM, China / 2019	
3B:217005195 (YLD & GN & BM)	3	-	10.19	IM, China / 2019	
3D:3034125 (YLD & GN)	3	2.22	11.64	IM, China / 2019	
6B:17675469	3	0.66	13.02	Cunderdin & Kalkee/2020	8.46 Mbp away from an MTA in (Li et al. 2019c)
7B:43893059	3	-	10.08	Cunderdin & Lock/2020	4.45 Mbp away from an MTA in (Juliana et al. 2019)

*MTAs with -log $_{10}$ (p) > 4 were considered as major MTAs.

**If MTAs showed significance in more than one location, the maximum R^2 is shown.

***Only those reported GRs with distance of less than 10 Mbp to the SNPs identified in this study are shown. MQTL names are according to a previous meta-QTL study (Liu et al. 2020).

SNPs with underlines indicate those that were considered as one locus on their individual chromosomes, because they were identified in one environment and were closely located (with a physical distance < 10 Mbp between each other).

SNP (chromosome: physical position)	Significance threshold* (-log$_{10}$ (p))	Additive effect	R^2 (%)**	Environment/year	Comparison with previously reported GRs***
7B:174805041	3	0.67	13.42	Brocklesby & Kalkee/2020	
*MTAs with -log$_{10}$ (p) > 4 were considered as major MTAs.					
**If MTAs showed significance in more than one location, the maximum R^2 is shown.					
***Only those reported GRs with distance of less than 10 Mbp to the SNPs identified in this study are shown. MQTL names are according to a previous meta-QTL study (Liu et al. 2020).					
SNPs with underlines indicate those that were considered as one locus on their individual chromosomes, because they were identified in one environment and were closely located (with a physical distance < 10 Mbp between each other).					

Although there were no overlapping MTAs between Australian and Chinese trials, some MTAs from the two trials were very close to each other, for example, on chromosome 1B, MTA 1B:215325944 identified in Australian trial was just 2.45 Mbp away from MTA 1B:212877972 identified in IM trial; on chromosome 2A, MTA 2A:1269018 in Australian trial was 5.24 Mbp away from 2A:6507322 in IM trial; on chromosome 2B, MTA 2B:287801761 in Australian trial was 5.96 Mbp away from an MTA 2B:293766172 responsible for TGW in IM trial; and on chromosome 3B, MTAs 3B:138977187 and 3B:205459416 in Australian trial were 5.14 and 3.09 Mbp away from 3B:133834102 and 3B:208548740 in IM trial, respectively (Fig. 2).

To identify major MTAs for other yield-related traits (PH, BM, GN, SN and TGW) across environments, a more stringent significant threshold of $-\log_{10}$ (p) > 4 were used. Due to the skews in the QQ plots for TGWs in the UWA Shenton Park trial and for SN in the IM trial (Fig. S3), the MTAs for these two traits were not considered to avoid false positive errors. Fifteen MTAs located on chromosome 1A, 1D, 2A, 6B and 7D were detected for PH, with two identified in the UWA Shenton Park trial and others in the IM trial; six MTAs located on chromosome 4A, 6A and 6B were detected for BM in the IM trial; four MTAs located on 1A, 1B and 3D were detected for GN with two each in the Shenton Park and IM trials; and two MTAs located on chromosome 2A and 2B for TGW in the IM trial. No overlaps among these markers were found, but noticeably one TGW marker SNP 2B:293766172 were found to be within the interval of a previously reported MQTL (MQTL2B.3) which had a physical position of 263220312−453816238 bp on chromosome 2B (Liu et al. 2020) (Table S4).

Candidate genes

Genes with functions involving sugar synthesis and transportation and stress responses were considered as important candidates for yield traits under rainfed conditions. Forty-six candidate genes were identified for the major or consistent MTAs (Table 4). Among them, two genes *TraesCS1A01G272200* and *TraesCS3B01G133400* harboured the SNPs within their introns, while others were closely located (within 2 Mbp) to their associated MTAs. Gene clusters containing consecutive genes of similar functions were also identified, such as *TraesCS1A01G271300* and *TraesCS1A01G271400* on chromosome 1A with a function of transmembrane proteins, *TraesCS2A01G002400* and *TraesCS2A01G002500* on chromosome 2A with a function of disease resistance TIR-NBS-LRR family, *TraesCS3B01G193700* and *TraesCS3B01G193800* on chromosome 3B with a function of zinc finger CCCH domain-containing protein 4, seven consecutive genes from *TraesCS3D01G008900* to *TraesCS3D01G009500* on chromosome 3D with a function of basic helix-loop-helix DNA-binding superfamily protein, and *TraesCS7A01G068200* to *TraesCS7A01G068300* on chromosome 7A with a function of NAC domain-containing protein.

Table 4
Candidate-genes for major marker-trait associations (MTAs) identified in this study.

Candidate genes*	Annotation	Gene-associated SNP	SNP-gene distance**
TraesCS1A01G271200, TraesCS1A01G271300 & TraesCS1A01G271400	Cyst nematode resistance protein-like; Transmembrane proteins (2 consecutive)	1A:465076391	751 bp away to first gene
TraesCS1A01G272200 & TraesCS1A01G272300	S-acyltransferase; Ethylene-responsive transcription factor	1A:465076391–465467894	SNP within intron of first gene
TraesCS1B01G147800	Myb	1B:212877972–215432253	1,023,997 bp away
TraesCS1B01G226600	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 3	1B:406281153	1,020,639 bp away
TraesCS2A01G002400 & TraesCS2A01G002500	Disease resistance protein (TIR-NBS-LRR class) family (2 consecutive genes)	2A:1269018	95,404 bp away
TraesCS2B01G154500	Transmembrane protein	2B:122702483	7645 bp away
TraesCS2B01G171900	Heat shock transcription factor	2B:144139063	2,301,035 bp away
TraesCS2B01G255300	E3 ubiquitin-protein ligase	2B:287801761	3,782,019 bp away
TraesCS2D01G080600	Myb/SANT-like DNA-binding domain protein	2D:34400629	23077 bp away
TraesCS3B01G133400	Alkaline alpha-galactosidase seed imbibition protein	3B:115783172	SNP within gene intron
TraesCS3B01G193300, TraesCS3B01G193700 & TraesCS3B01G193800	Zinc finger CCCH domain-containing protein 4 (gene cluster with the latter 2 consecutive)	3B:208548740	811,684 bp away
TraesCS3B01G194900	E3 ubiquitin-protein ligase	3B:217005195	1,386,390 bp away
TraesCS3D01G008900 - TraesCS3D01G009500	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein (gene cluster with 7 consecutive genes)	3D:3034125	6,632 bp away
TraesCS3D01G169100 & TraesCS3D01G169200	23S rRNA (uracil(747)-C(5))-methyltransferase RlmC	3D:145929682	728,010 bp away
TraesCS3D01G241200	RADIALIS-like transcription factor (Myb like)	3D:334295007	132,037 bp away
TraesCS4B01G147700	ATPase subunit 4	4B:209651192	339,516 bp awy
TraesCS5A01G116000	Chaperone DnaJ-domain superfamily protein	5A:237084603	285,414 bp away

*Those with more than one candidates indicate they are in gene clusters.

**For those with more than one marker or gene, the closest marker-gene distance is shown.

Candidate genes*	Annotation	Gene-associated SNP	SNP-gene distance**
TraesCS5A01G119300	Histone deacetylase family protein, expressed	5A:248358792	173,841 bp away
TraesCS5A01G036900	NBS-LRR-like resistance protein	5A:29297520–33961012	34,684 bp away
TraesCS5B01G101900	RNA polymerase-associated RTF1-like protein	5B:135262608	333,942 bp away
TraesCS5D01G089100	Formin-like protein	5D:95706398	153,881 bp away
TraesCS6B01G029400, TraesCS6B01G029800 & TraesCS6B01G029900	NBS-LRR resistance-like protein	6B:17675469	51,005 bp away
TraesCS7A01G068000 - TraesCS7A01G068300	NAC domain-containing protein (gene cluster); Transcription elongation factor 1	7A:34077500	25,800 bp away
TraesCS7B01G044300	Ubiquitin	7B:43893059	55,856 bp away
TraesCS7B01G129800	Deaminase-related family protein	7B:155982479	74,965 bp away
TraesCS7B01G138900	Auxin response factor	7B:174805041–176068517	229,919 bp away
TraesCS7B01G162500 & TraesCS7B01G162600	Receptor protein kinase; Stress up-regulated Nod 19 protein	7B:224008352	829,531 bp away
*Those with more than one candidates indicate they are in gene clusters.			
**For those with more than one marker or gene, the closest marker-gene distance is shown.			

Maturity vs. yield

The reported maturity MTAs were shown in Fig. S4. They densified the most on chromosome 5B at the interval of 403782889 to 681563146 bp, harbouring 60 maturity MTAs for all environments including irrigated, drought stressed, and heat stressed. Many MTAs for yield traits were also identified in this area (Fig. 1), suggesting that these GRs regulate yield traits in correlation to maturity for all environments. The other densified segments for maturity MTAs were found on chromosome 1A, 2B, 6A, 6B and 7D, with chromosome 6B had the most maturity MTAs for drought or heat stressed environments. This suggested that yield GRs overlapping with these areas need to be considered for their correlations with maturity under the specific individual environments.

Discussion

Compared to other crops, there is still significant room for improvement of yield in wheat. Breeding robust varieties by pyramiding favourable alleles will allow crops to perform better in target environments and even grow in marginal areas that are currently not effectively used. This study focuses on rainfed conditions to compare significant GRs for wheat yield under different environments with similar rainfed farming practices. A comparison of them with previously reported GRs identified by GWAS and MQTL analyses is also done to see if the results in this study are consistent with other studies.

The availability of the wheat reference genome sequence makes it possible to compare different studies at physical position level. The mini-review of GWAS has revealed that the reported MTAs/GRs responsible for yield are widely distributed throughout the 21 wheat chromosomes. An interesting phenomenon is that many of the MTAs/GRs tend to

densely located together as clusters at certain parts of the chromosome segments with possible physical lengths as long as 100 Mbp (Fig. 1). As expected, many of these segments are located towards the telomere ends of chromosomes where are known as gene-rich regions. However, some segments located towards the middle parts of the chromosomes also contain a large number of MTAs — such as those on chromosome 3B, 5A, 5B, 6B and 6D, making them "hotspot" segments for harbouring yield associated markers other than the chromosome ends. Among all these "hotspot" segments, some have been identified in both QTL linkage mapping and GWAS, such as the segments on chromosome 2A, 2B, 4B, 5A, 6B, 6D, 7A and 7B, each overlapping with one or more MQTL, suggesting that they play important and consistent roles to regulate yield performance under different environments. Further looking at the distribution of the newly identified MTAs/loci in this study, it can also be observed that the MTAs/loci mostly distributed as clusters at certain chromosome segments. Although no overlapping MTAs are found between Australian and Chinese trials, many of the markers from the two trials are physically very close to each other showing clustering at "hotspot" chromosome segments. To avoid high-LD-related false positive errors, we have used a stringent pairwise correlation cut-off value of $r^2 > 0.8$, therefore the MTA clusters observed in this study may not be simply due to the LD between the markers, especially when those clustered MTAs are identified from different environments. The existence of such MTA clustered chromosome segments is further proved by the GWAS mini-review in which MTA clusters are also observed where many yield-associated MTAs from different studies under different environments also tend to densely distributed at certain chromosome segments. These "hotspot" chromosome segments are mostly located on A and B subgenomes, whereas on D subgenome the markers may be in clusters but with much less densities, such as those segments on chromosome 2D, 3D, and 6D, suggesting the A and B subgenomes are the major players in controlling yield-related traits.

Some of the MTAs or loci newly identified in this study fall in the "hotspot" chromosome segments harbouring many previously reported MTAs/GRs, while others show far distance from previously reported ones but mostly in clusters at certain chromosome segments, such as those on chromosome 1A, 3B, 4B, 5B and 7A (Fig. 1). These newly identified "hotspot" chromosome segments could be specifically responsible for yield traits under rainfed farming conditions similar to Australia or North China. When searching for candidate genes, many gene clusters are found to associate with the major or consistent MTAs. Gene clusters have been previously reported to frequently appear in yield-associated MQTL regions (Liu et al. 2020). How gene clusters work together, for example, if they work in sequence under the same environment or if they work separately under different environments, is still unclear (Medema et al. 2015).

We hypothesize that the existence of the "hotspot" chromosome segments may work as a bigger version of gene clusters for controlling yield traits which are largely dependent on individual environments. Within the "hotspot" segment, there may be a large number of closely located genes, with similar functions or with epistatic effects to each other, that may be selectively active or expressed under different environments. Although different gene networks are triggered in different environments, they have similar effects towards yield-related traits, therefore such similar genes are clustered together at certain segments of the chromosome. This could be the reason for the existence of gene clusters as well. The implication of these findings to practical breeding is as suggested by Liu et al. (Liu et al. 2020) that marker assisted selection (MAS) or transgenic method targeting a single gene might not be as effective as MAS targeting a much larger GR where all the genes or gene clusters underlying the GR play important roles.

Wheat yield is a polygenic trait controlled by many genes with small effects which can significantly vary according to environments. The large and complex wheat genome, and the narrow adaptation of the crop due to long history of domestication and inbreeding further impede the progress for yield breeding. The increased stress conditions due to climate change in recent years have made an urgent call for efficient yield improvement. Genomic selection using DNA sequence based markers provides opportunities to tackle the issue. For genomic selection, clearly there is a need to make a balance between the cost and efficiency of the method. Julianna et al. (Juliana et al. 2019) investigated the effects of different subsets of markers filtered according to different criteria on GWAS and reported that genomic coverage had minimal impact on genomic prediction of traits once marker density has reached to a certain genomic resolution. They found that filtered GBS markers (ranged from 9k - 16k in number) with a pairwise correlations from less than 0.8 to less than 0.5 achieved

similar prediction accuracy. In our study, after filtering with stringent selection criteria, we only used 15% of the clean sequence data for the GWAS analysis. We suggest that the cost of genomic selection can be reduced significantly if a more efficient selection of the DNA sequence based markers can be achieved. Therefore, the "hotspot" chromosome segment and gene cluster regions could be the major GRs to target for genomic selection of yield-related traits in wheat.

Declarations

Funding

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data availability

The ddRAD sequencing data for the Australian and Chinese populations are deposited in the China National Genomics Data Center under accession numbers. The phenotypic data are available in Supplementary Table S3.

Code availability

Not applicable

Author contributions

HL, DM, SZ, YZ, YW, AZ, ZL and GY conceived and designed the experiment, who were investigators of the Global Innovation Linkages Project. HL, DM, YZ, AZ, ZL and YW collected the association panel seeds. HL, DM, YJ, XZ, GL, WY and GY conducted the field experiments. HL, SZ, CZ and KC analysed the data, and HL wrote the manuscript. All authors critically reviewed the manuscript.

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

References

1. Acuna-Galindo MA, Mason RE, Subramanian NK, Hays DB (2015) Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. Crop Sci 55(2):477–492
2. Ain Q-u, Rasheed A, Anwar A, Mahmood T, Mahmood T, Imtiaz M, He Z, Xia X, Quraishi U (2015) Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. Front Plant Sci 6:743

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

4. Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246

5. Alqudah AM, Sallam A, Stephen Baenziger P, Börner A (2020) GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review. J Adv Res 22:119–135

6. Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Stein N, Choulet F, Distelfeld A, Poland J, Ronen G, Barad O, Stein N, Barad O, Mascher M, Ben-Zvi G, Sharpe AG, Ben-Zvi G, Balfourier F, Rogers J, Hayden M, Koh CS, Josselin A-A, Koh C, Paux E, Rigault P, Pozniak CJ, Sharpe AG, Tibbits J, Rogers J, Choulet F, Lang D, Gundlach H, Keeble-Gagnère G, Mayer KFX, Wicker T, Prade V, Rimbert H, Wicker T, Guilhot N, Rimbert H, Felder M, Leroy P, Kaithakottil G, Lang D, Leroy P, Lux T, Abrouk M, Appels R, Uauy C, Appels R, Fischer I (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:661

7. Bhatta M, Morgounov A, Belamkar V, Baenziger PS (2018) Genome-wide association study reveals novel genomic regions for grain yield and yield-related rraits in drought-stressed synthetic hexaploid wheat. Int J Mol Sci 19(10):3011

8. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

9. Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-generation reference panels. Am J Hum Genet 103:338–348

10. Cumer T, Pouchon C, Boyer F, Yannic G, Rioux D, Bonin A, Capblancq T (2021) Double-digest RAD-sequencing: do pre- and post-sequencing protocol parameters impact biological results? Mol Genet Genomics 296:457–471

11. Darrier B, Russell J, Milner SG, Hedley PE, Shaw PD, Macaulay M, Ramsay LD, Halpin C, Mascher M, Fleury DL, Langridge P, Stein N, Waugh R (2019) A comparison of mainstream genotyping platforms for the evaluation and use of barley genetic resources. Front Plant Sci 10:544

12. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498

13. Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929

14. Gahlaut V, Jaiswal V, Singh S, Balyan HS, Gupta PK (2019) Multi-locus genome wide association mapping for yield and its contributing traits in hexaploid wheat under different water regimes. Sci Rep 9:19486

15. Garcia M, Eckermann P, Haefele S, Satija S, Sznajder B, Timmins A, Baumann U, Wolters P, Mather DE, Fleury D (2019) Genome-wide association mapping of grain yield in a diverse collection of spring wheat (Triticum aestivum L.) evaluated in southern Australia. PLOS ONE 14:e0211730

16. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J (2020) RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Computer Science 6:e251

17. He W, Zhao S, Liu X, Dong S, Lv J, Liu D, Wang J, Wang J, Meng Z (2013) ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis. Genetics molecular research 12:6275–6283

18. Jamil M, Ali A, Gul A, Ghafoor A, Napar AA, Ibrahim AMH, Naveed NH, Yasin NA, Mujeeb-Kazi A (2019) Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. BMC Plant Biol 19:149

19. Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U, Bhavani S, Singh PK, Randhawa MS, He X, Guzman C, Dreisigacker S, Rouse MN, Jin Y, Pérez-Rodríguez P, Montesinos-López OA, Singh D, Mokhlesur Rahman M, Marza F, Singh RP (2019) Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. Nat Genet 51:1530–1539

20. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549

21. Li F, Wen W, Liu J, Zhang Y, Cao S, He Z, Rasheed A, Jin H, Zhang C, Yan J, Zhang P, Wan Y, Xia X (2019a) Genetic architecture of grain yield in bread wheat based on genome-wide association studies. BMC Plant Biol 19:168

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

23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

24. Li L, Mao X, Wang J, Chang X, Reynolds M, Jing R (2019b) Genetic dissection of drought and heat-responsive agronomic traits in wheat. Plant Cell Environment 42:2540–2553

25. Li L, Peng Z, Mao X, Wang J, Chang X, Reynolds M, Jing R (2019c) Genome-wide association study reveals genomic regions controlling root and shoot traits at late growth stages in wheat. Ann Bot 124:993–1006

26. Li X, Xu X, Liu W, Li X, Yang X, Ru Z, Li L (2020) Dissection of superior alleles for yield-related traits and their distribution in important cultivars of wheat by association mapping. Front Plant Sci 11:175

27. Liu H, Mullan D, Zhang C, Zhao S, Li X, Zhang A, Lu Z, Wang Y, Yan G (2020) Major genomic regions responsible for wheat yield and its components as revealed by meta-QTL and genotype-phenotype association analyses. Planta 252:65

28. Liu J, Feng B, Xu Z, Fan X, Jiang F, Jin X, Cao J, Wang F, Liu Q, Yang L, Wang T (2017) A genome-wide association study of wheat yield and quality-related traits in southwest China. Mol Breeding 38:1

29. Liu J, Xu Z, Fan X, Zhou Q, Cao J, Wang F, Ji G, Yang L, Feng B, Wang T (2018) A Genome-Wide Association Study of Wheat Spike Related Traits in China. Frontiers in Plant Science 9

30. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng Y-Q, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJM, Ehling-Schulz M, Entian K-D, Fischbach MA, Gerwick L, Gerwick WH, Gross H, Gust B, Hertweck C, Höfte M, Jensen SE, Ju J, Katz L, Kaysser L, Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon H-J, Lautru S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Méndez C, Metsä-Ketelä M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Müller R, Neilan BA, Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM, Pernodet J-L, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B, Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Süssmuth RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T, van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C, Zotchev SB, Breitling R, Takano E, Glöckner FO (2015) Minimum information about a biosynthetic gene cluster. Nature Chemical Biology 11:625–631

31. Muhu-Din Ahmed HG, Sajjad M, Zeng Y, Iqbal M, Habibullah Khan S, Ullah A, Nadeem Akhtar M (2020) Genome-wide association mapping through 90K SNP array for quality and yield attributes in bread wheat against water-deficit conditions. Agriculture 10:392

32. Pang Y, Liu C, Wang D, St. Amand P, Bernardo A, Li W, He F, Li L, Wang L, Yuan X, Dong L, Su Y, Zhang H, Zhao M, Liang Y, Jia H, Shen X, Lu Y, Jiang H, Wu Y, Li A, Wang H, Kong L, Bai G, Liu S (2020) High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol Plant 13:1311–1327

33. 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. The American Journal of Human Genetics 81:559–575

34. Qaseem MF, Qureshi R, Shaheen H, Shafqat N (2019) Genome-wide association analyses for yield and yield-related traits in bread wheat (Triticum aestivum L.) under pre-anthesis combined heat and drought stress in field conditions. PLOS ONE 14:e0213407

35. Sehgal D, Autrique E, Singh R, Ellis M, Singh S, Dreisigacker S (2017) Identification of genomic regions for grain yield and yield stability and their epistatic interactions. Sci Rep 7:41578

36. Sehgal D, Rosyara U, Mondal S, Singh R, Poland J, Dreisigacker S (2020) Incorporating genome-wide association mapping results into genomic prediction models for grain yield and yield stability in CIMMYT spring bread wheat. Front Plant Sci 11:197

37. Sharma R, Cockram J, Gardner KA, Russell J, Ramsay L, Thomas WTB, O'Sullivan DM, Powell W, Mackay IJ (2020) Trends of genetic changes uncovered by Env- and Eigen-GWAS in wheat and barley. bioRxiv:2020.2011.2027.400333

38. Sonah H, O'Donoughue L, Cober E, Rajcan I, Belzile F (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol J 13:211–221

39. Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP (2015) Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor Appl Genet 128:353–363

40. Sun C, Zhang F, Yan X, Zhang X, Dong Z, Cui D, Chen F (2017) Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. Plant Biotechnol J 15:953–969

41. Tsai H-Y, Janss LL, Andersen JR, Orabi J, Jensen JD, Jahoor A, Jensen J (2020) Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. Sci Rep 10:3347

42. Wang X, Liu H, Liu G, Mia MS, Siddique KHM, Yan G (2019) Phenotypic and genotypic characterization of near-isogenic lines targeting a major 4BL QTL responsible for pre-harvest sprouting in wheat. BMC Plant Biol 19:348

43. Ward BP, Brown-Guedira G, Kolb FL, Van Sanford DA, Tyagi P, Sneller CH, Griffey CA (2019) Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. PLOS ONE 14:e0208217

44. Wu Y, Zhou Z, Dong C, Chen J, Ding J, Zhang X, Mu C, Chen Y, Li X, Li H, Han Y, Wang R, Sun X, Li J, Dai X, Song W, Chen W, Wu J (2020) Linkage mapping and genome-wide association study reveals conservative QTL and candidate genes for Fusarium rot resistance in maize. BMC Genom 21:357

45. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X (2021) rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics Proteomics Bioinformatics. doi:10.1016/j.gpb.2020.10.007

46. Zhang LY, Liu DC, Guo XL, Yang WL, Sun JZ, Wang DW, Zhang A (2010) Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. J Integr Plant Biol 52:996–1007

Figures
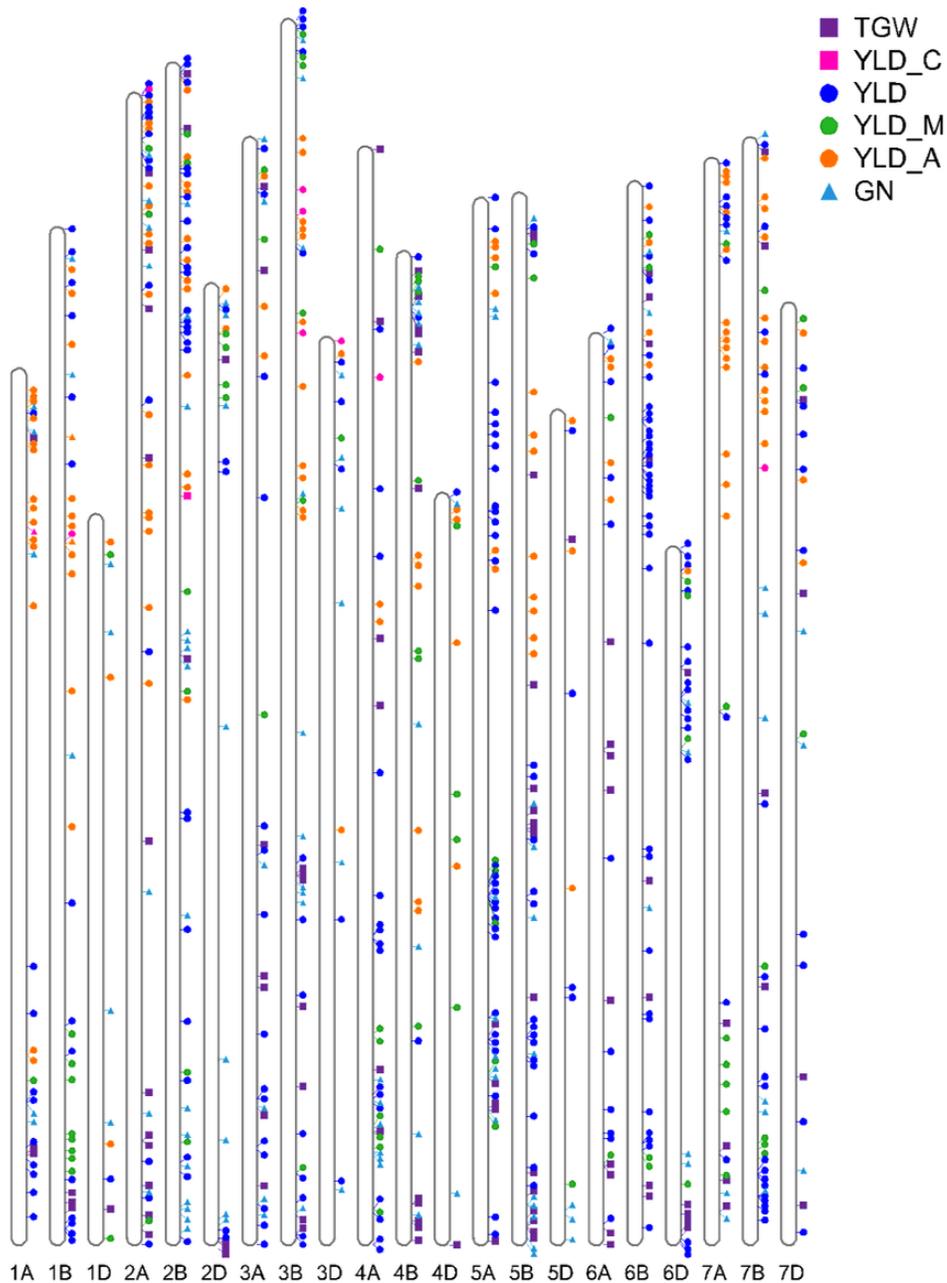
Figure 1

Genomic regions (GRs) responsible for yield-related traits identified in previous GWAS (labels TGW, YLD and GN) (Ain et al. 2015; Akram et al. 2021; Bhatta et al. 2018; Gahlaut et al. 2019; Garcia et al. 2019; Jamil et al. 2019; Juliana et al. 2019; Li et al. 2019a; Li et al. 2019b; Li et al. 2019c; Li et al. 2020; Liu et al. 2017; Liu et al. 2018; Muhu-Din Ahmed et al. 2020; Pang et al. 2020; Qaseem et al. 2019; Sehgal et al. 2017; Sehgal et al. 2020; Sharma et al. 2020; Sukumaran et al. 2015; Sun et al. 2017; Tsai et al. 2020; Ward et al. 2019), previous meta-QTL analysis (label YLD_M) (Liu et al. 2020), and this study (labels YLD-A and YLD_C). Previously reported GRs were labelled in cold colours including blue, green and purple, whereas GRs identified in this study were labelled in warm colours including pink and orange. YLD_A: yield in Australian trial; YLD_C: yield in Chinese trial. Bar represents 100Mb physical distance.
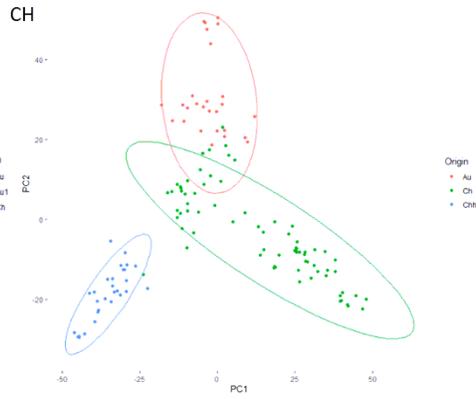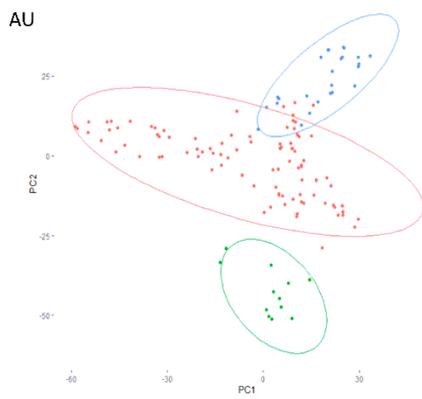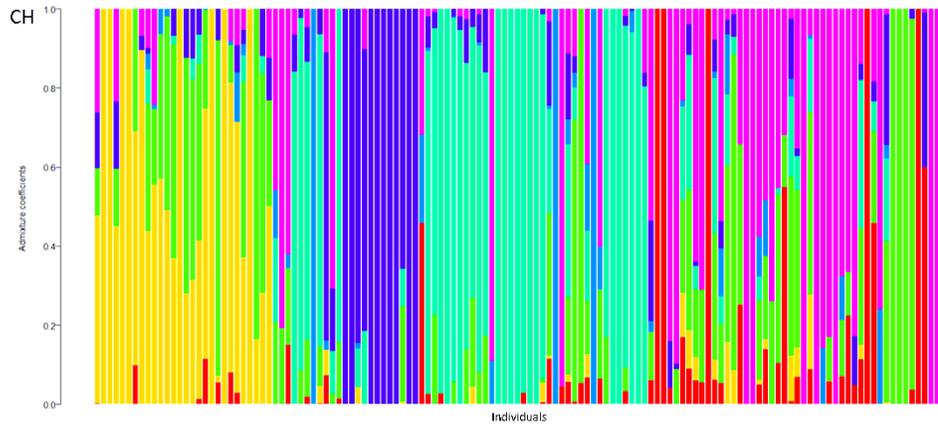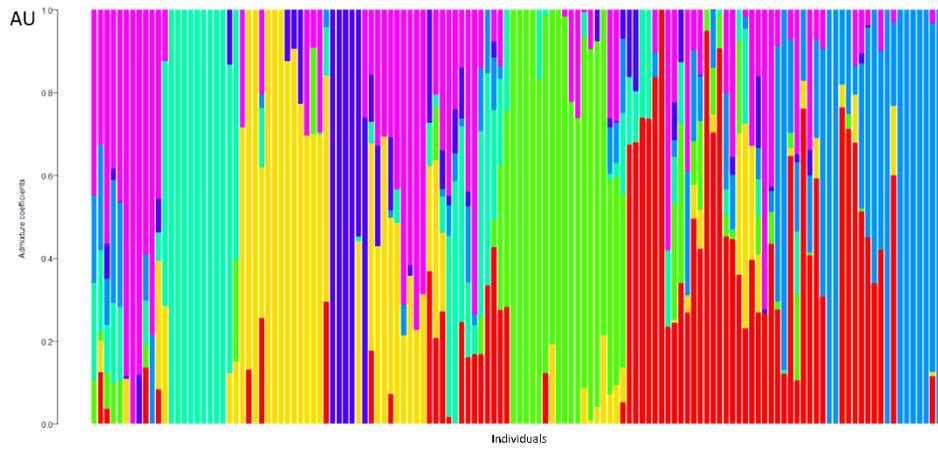
Figure 2

Population structures of the varieties investigated in the Australian (AU) and the Chinese (CH) trials, as revealed by structure analyses (top two) and principal component analyses (bottom).
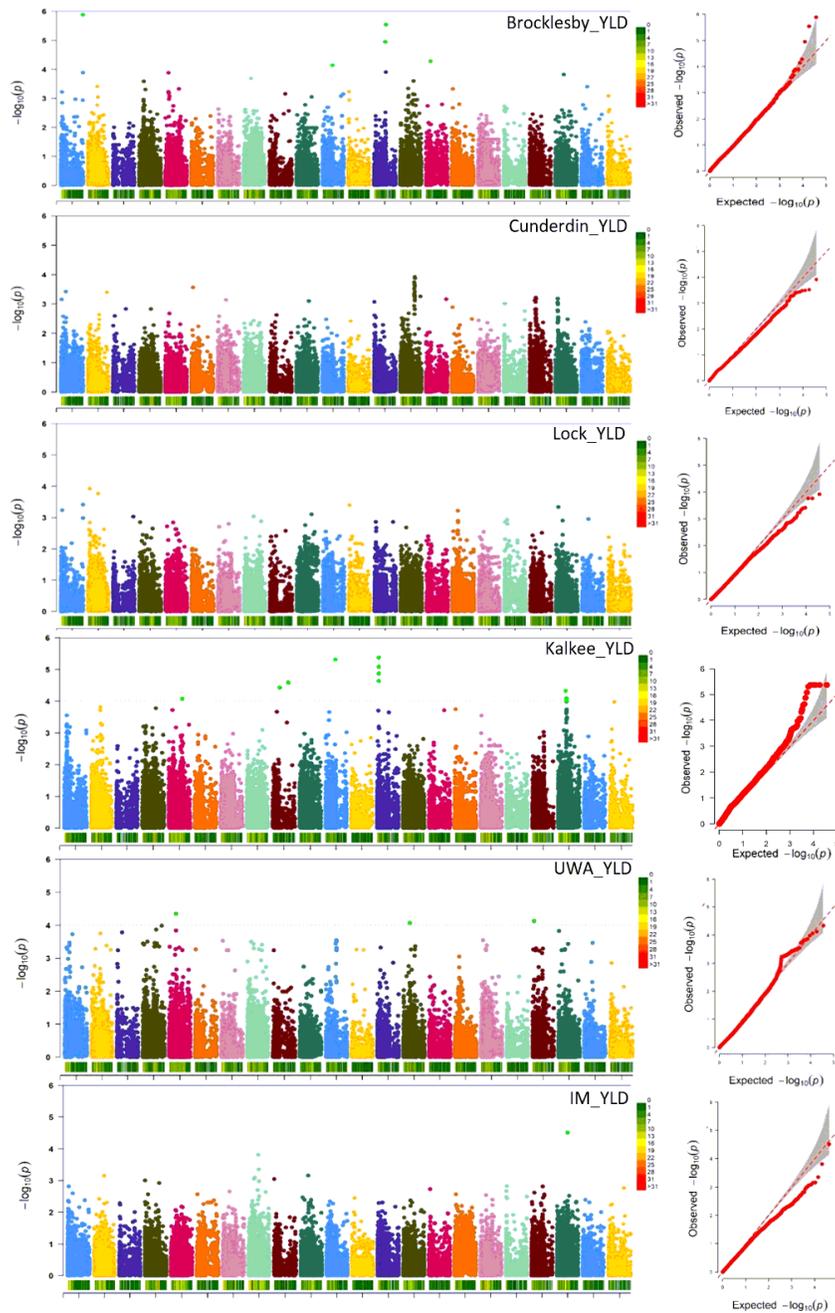
Figure 3

Manhattan and QQ plots of GWAS on yield in spring wheat. X-axis shows the chromosome position while chromosome Unk represents unmapped SNPs; and Y-axis is P-value (−log10). The horizontal lines in the Manhattan plots indicate threshold for significance at -log10 (p) > 4 or 6. Those marker-trait associations (MTAs) with -log10 (p) > 4 are shown in light green dots. UWA, UWA Shenton Park Field Station; IM: Inner Mongolia.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalinformation.pdf