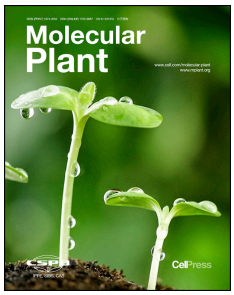
Journal Pre-proof



Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement

Jizeng Jia, Guangyao Zhao, Danping Li, Kai Wang, Chuizheng Kong, Pingchuan Deng, Xueqing Yan, Xueyong Zhang, Zefu Lu, Shujuan Xu, Yuannian Jiao, Kang Chong, Xu Liu, Dangqun Cui, Guangwei Li, Yijing Zhang, Chunguang Du, Liang Wu, Tianbao Li, Dong Yan, Kehui Zhan, Feng Chen, Zhiyong Wang, Lichao Zhang, Xiuying Kong, Zhengang Ru, Daowen Wang, Lifeng Gao

PII: S1674-2052(23)00329-5

DOI: https://doi.org/10.1016/j.molp.2023.10.015

Reference: MOLP 1632

To appear in: *MOLECULAR PLANT*

Received Date: 14 August 2022

Revised Date: 12 July 2023

Accepted Date: 23 October 2023

Please cite this article as: Jia J., Zhao G., Li D., Wang K., Kong C., Deng P., Yan X., Zhang X., Lu Z., Xu S., Jiao Y., Chong K., Liu X., Cui D., Li G., Zhang Y., Du C., Wu L., Li T., Yan D., Zhan K., Chen F., Wang Z., Zhang L., Kong X., Ru Z., Wang D., and Gao L. (2023). Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement. Mol. Plant. doi: https://doi.org/10.1016/j.molp.2023.10.015.

Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement

Jizeng Jia[1,2,12], Guangyao Zhao[2,12], Danping Li[2,12], Kai Wang[4,12], Chuizheng Kong[2,12], Pingchuan Deng[5,11], Xueqing Yan[6,7], Xueyong Zhang[2], Zefu Lu[2], Shujuan Xu[7,8], Yuannian Jiao[6,7], Kang Chong[7,8], Xu Liu[2], Dangqun Cui[1], Guangwei Li[1], Yijing Zhang[9], Chunguang Du[1], Liang Wu[5,10], Tianbao Li[1,2], Dong Yan[2], Kehui Zhan[1], Feng Chen[1], Zhiyong Wang[1], Lichao Zhang[2], Xiuying Kong[2,*], Zhengang Ru[3,*], Daowen Wang[1,*], Lifeng Gao[2,*]

[1]College of Agronomy, Collaborative Innovation Center of Henan Grain Crops, State Key Laboratory of Wheat and Maize Crop Science, and Center for Crop Genome Engineering, Henan Agricultural University, Zhengzhou 450046, Henan, China

[2]State Key Laboratory of Crop Gene Resources and Breeding, the National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[3]School of Life Science and Technology, Henan Institute of Science and Technology, Xinxiang 453003, Henan, China

[4]Xi'An Shansheng Biosciences Co., Ltd., Xi'an 710000, China

[5]Zhejiang Provincial Key Laboratory of Crop Genetic Resources, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, Zhejiang, China

[6]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[7]University of Chinese Academy of Sciences, Beijing 100049, China

[8]Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[9]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, China

[10]Hainan Yazhou Bay Seed Laboratory, Hainan Institute of Zhejiang University, Sanya 562000, Hainan, China

[11]State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling 612100, Shaanxi, China

[12]These authors contributed equally.

[*]**Correspondence:** Xiuying Kong (kongxiuying@caas.cn), Zhengang Ru (rzgh58@163.com), Daowen Wang (dwwang@henau.edu.cn), Lifeng Gao (gaolifeng@caas.cn)

Short summary

Genome resources generated for an elite wheat cultivar are analyzed, which has yielded new insights into the genomic changes in recent varietal improvement and subgenome diploidization and divergence in common wheat. This leads to the development of a homoeologous locus-based GWAS approach highly effective for unraveling the agronomic trait-associated loci and their superior haplotypes valuable for genomics-assisted breeding.

ABSTRACT

Despite the progress made recently in crop genomics studies, the genomic changes brought by modern breeding selection are still poorly understood, thus hampering genomics-assisted breeding especially in the polyploid crops with compound genomes such as common wheat (*Triticum aestivum*). In this work, we constructed genome resources for the modern elite common wheat variety Aikang 58 (AK58). Comparisons between AK58 and the landrace cultivar Chinese Spring (CS) shed light on genomic changes occurred in recent wheat varietal improvement. Furthermore, we explored subgenome diploidization and divergence in common wheat and developed a homoeologous locus-based GWAS (HGWAS) approach, which was more effective than single homoeolog-based GWAS in unraveling agronomic trait-associated loci. A total of 123 major HGWAS loci were detected using the genetic population derived from AK58 and CS. Elite homoeologous haplotypes (HHs), formed by combinations of subgenomic homoeologs of the associated loci, were found in both parents and progenies, many of which could substantially improve wheat yield and related traits. We builded a website (available in https://triticeae.henau.edu.cn/aikang58/) in which data download of AK58 genome assembly sequence and annotation, blast analysis and Jbrowse could be performed. Our work enriches wheat genome resources, provides new insight into the genomic changes involved in modern wheat improvement, and suggests that efficient mining of elite HHs may contribute substantially to genomics-assisted breeding in common wheat and other polyploid crops.

Key words: common wheat, genome sequencing, subgenome diploidization and divergence, homoeologous locus-based GWAS, homoeologous haplotypes, polyploid crops

INTRODUCTION

Since the completion of *Arabidopsis* and rice genome sequence, dissecting and improving plant traits have entered the genomics era (Sun et al., 2022). The availability of genome sequences and high genome coverage molecular markers, e.g., single nucleotide polymorphism (SNP) markers, has greatly increased the efficiency of isolating agronomically important genes either by forward and reverse genetic research or using genome-wide association study (GWAS) (Gupta et al., 2019; Tibbs Cortes et al., 2021). This has led to a large burst of functionally characterized genes and thus improved understanding of the genetic architecture of many agronomic traits (Soyk et al., 2020). Aided by multi-omics and genome-editing technologies, the molecular mechanisms operating in trait formation are also being revealed at a fast pace (Weckwerth et al., 2020; Gao, 2021; Scossa et al., 2021). Together, these achievements are paving the way for genomics-assisted crop improvement especially in the crops with simpler diploid genomes such as rice and maize (Purugganan and Jackson, 2021; Varshney et al., 2021). However, as the majority of plant traits are complex and controlled by polygenes and affected by environmental conditions, prodigious efforts are still needed to deepen understanding of the genomic and molecular basis of agronomic traits. This is particularly relevant for the crops with large and complex polyploid genomes (Michael and Van Buren, 2015; May et al., 2023).

Up to 70% of the flowering plants on earth may be recent polyploids, and approximately 40% - 50% of the cultivated crops have polyploid genomes (Wood et al., 2009; Moghe and Shiu, 2014; Salman-Minkov et al., 2016). Many important food, fiber, and oil crops, such as common wheat (*Triticum aestivum*, AABBDD, 2n = 6x = 42), upland cotton (*Gossypium hirsutum*, AADD, 2n = 4x = 52), and peanut (*Arachis hypogaea*, AABB, 2n = 4x = 40) are allopolyploids carrying two or more related but not identical subgenomes (Song et al., 2017; Zhuang et al., 2019). Compared to

diploid crops (e.g., rice and maize), each typical homoeologous locus in an allopolyploid include at least two subgenomic homoeologs, whose biological functions are frequently affected by non-functionalization, subfunctionalization, or neofunctionalization (Lynch and Conery, 2000; Ma and Gustafson, 2005; Jackson and Chen, 2010). Furthermore, unique to polyploids, parental homoeologs are subjected to reassortment in the offspring in hybridization breeding, which can generate many alternative combinations of subgenomic homoeologs. For example, in a biparental F_2 population of an allohexaploid crop such as common wheat, for each typical triad locus with three subgenomic orthologs, reassortment of allelic parental homoeologs will yield 27 (3^3) combinations of homoeologs, of which two are parental and 25 are newly formed. Some of these homoeolog combinations may confer improved traits than the corresponding locus of the better parent, and thus representing elite homoeologous haplotypes (HHs) valuable for enhancing both wheat genetic diversity and trait performance. But this aspect has seldom been investigated in depth and conscientiously exploited in polyploid crop improvement in the past. This is mainly caused by the lack of an efficient approach for detecting agronomically important homoeologous loci at genome-wide level in polyploid crops. The vast amount of GWAS investigations published to date for polyploid plants generally use one homoeolog from a single subgenome, rather than all homoeologs, in genotyping and association tests. Consequently, an alternative GWAS approach, which utilizes the molecular variations of all homoeologs as genotyping information to detect trait controlling genes, is needed, which will largely facilitate the mining and exploration of HHs in common wheat and other polyploids.

Common wheat is not only a major staple crop in the world but also a model for studying the unique genome biology of polyploid plants (Dubcovsky and Dvorak, 2007; Venske et al., 2019). Its A and D subgenomes were donated by *T. urartu* (A^UA^U, 2n = 2x = 14) and *Aegilops tauschii* ($D^{Aet}D^{Aet}$, 2n = 2x = 14), respectively, while the B subgenome might be derived from an unidentified species related to *Ae. speltoides* (Levy and Feldman, 2022). Two polyploidization events occurred in the formation of

hexaploid wheat. The first one took place around 0.8 million years ago and gave rise to tetraploid wild emmer wheat (WEW, *T. turgidum* ssp. *dicoccoides*, AABBDD, 2n = 4x = 28); the second one happened about 10,000 years ago and yielded the ancestral hexaploid wheat, which subsequently diverged into different cultivated forms, with common wheat becoming the most widely cultivated food crop and accounting for over 90% of the global wheat production today (Shewry and Hey, 2015; Levy and Feldman, 2022). As uncovered by recent genomic analysis, hexaploid wheat has dispersed origin and protracted speciation and domestication history, with frequent interploidy introgressions playing a prominent role in shaping its polyploid genome (Zhou et al., 2020; Wang et al., 2022a; Zhao et al., 2023b). Owing to intensive selection, modern breeding has decreased the genetic diversity of common wheat, but introduced a number of alien chromatins from wheat relatives, which enhance wheat yield potential particularly in the environments with high biotic and/or abiotic stresses (Walkowiak et al., 2020; Przewieslik-Allen et al., 2021). Thus, the composition and function of modern common wheat genomes are highly dynamic and plastic, which enables them to adapt to contrasting environments and to produce the grains with different end-use requirements. Clearly, only one reference genome sequence based on the landrace cultivar Chinese Spring (CS) is not sufficient to cover the global diversities of common wheat (IWGSC et al., 2018); pangenomic resources, as well as the genome databases generated from using regionally important elite cultivars, are needed to aid genomics-assisted breeding in common wheat (Walkowiak et al., 2020). Consequently, the genomes of several elite common wheat varieties from American, Asian, and European countries have been sequenced and analyzed recently (Sato et al., 2021; Shimizu et al., 2021; Akpinar et al., 2022; Athiyannan et al., 2022; Aury et al., 2022; Kale et al., 2022; Shi et al., 2022).

With a yearly planting area over 24 million hectares and an annual production exceeding 130 million tons, China is the world's largest wheat producer and consumer (Xiao et al., 2022). Among the ten wheat cultivation zones of China, the Yellow and Huai River Valley (YHRV) winter wheat region is most important, as it contributes

over 70% to the total national annual wheat production (He et al., 2014). Aikang 58 (AK58), a leading elite winter wheat cultivar in the YHRV region since its release in 2005, exhibits strong resistance to lodging and elevated tolerance to multiple abiotic stresses (e.g., drought and frost) (Wang et al., 2018; Jia et al., 2021). It carries the beneficial 1RS translocation and favorable genes in vernalization (i.e., *vrn-A1*, *vrn-B1*, *vrn-B3*, *vrn-D1*), plant height (e.g., *Rht1*), photoperiod control (e.g., *Ppd-D1a*), and end-use quality (e.g., *Glu-D1d*) traits (http://wheatpedigree.net/sort/show/111306). Furthermore, AK58 has been used as a key parental germplasm for developing more than 100 commercial common wheat cultivars in China (Wang et al., 2018; Jia et al., 2021). Therefore, AK58 is a typical product of intensive selection breeding and a valuable genetic resource for further wheat improvement. Studying the genomic and molecular basis underlying AK58's outstanding performance may shed new light on the changes conferred by modern breeding as well as generate novel resources for future wheat improvement in the genomics era. Thus, we developed a comprehensive genome database for AK58 and initiated a series of genetic and breeding studies using AK58's genome database.

Previously, we reported the 3D genome characteristics, the distribution and evolutionary significance of Helitron transposons, and the centromere structures in AK58 (Jia et al., 2021; Wang et al., 2022b). The main objectives of this work were to outline in detail the major components of AK58's genome database and its application in revealing the genomic changes involved in modern wheat improvement. Furthermore, we explored subgenome diploidization and divergence and designed a homoeologous locus-based GWAS (HGWAS) to identify the polyploid loci functioning in agronomic trait control. In this approach, the three subgenomic homoeologs of a homoeologous locus were each tagged by a closely linked SNP marker, which facilitated the distinguishment of both parental and progeny HHs. Subsequently, these HHs were used as genotyping data for GWAS computation. It is worth noting that this approach could also be employed for HGWAS analysis using natural varietal populations after the HHs of homoeologous loci were identified using

commercial SNP arrays. We found that HGWAS was more effective for uncovering the loci controlling important crop traits than the conventional single homoeolog-based GWAS in common wheat, and unraveled 123 homeologous loci highly significantly associated with the examined agronomic traits using the genetic population derived from AK58 and CS, a model landrace with a well sequenced genome (IWGSC et al., 2018). Remarkably, many of the HHs mined in this work could largely improve wheat yield and related traits, thus having the potential to accelerate wheat improvement through genomics-assisted breeding.

RESULTS

Chromosome-scale assembly of AK58 genome

By using the sequence data generated from Illumina sequencing, PacBio single molecule real time sequencing, 10× Genomics linked reads, and Hi-C analysis, pseudomolecule sequences representing the 21 chromosomes of AK58 genome were assembled (see Methods). The assembly consisted of 279,861 contigs (N50, 237.2 kb) and 159,139 scaffolds (N50, 28.3 Mb) (Table 1). After integrating Hi-C data, the scaffold N50 was increased to 715 Mb (Supplemental Table 1). The total scaffold length of AK58 assembly (14.75 Gb) spanned 95.2% of the 15.5 Gb estimated genome size of common wheat (IWGSC et al., 2018), with the top 584 scaffolds covering 90% of the assembly (Table 1). Through combining Hi-C information and a high-density genetic map, nearly 97% of the assembled sequences were anchored to and ordered on the 21 pseudochromosomes (Supplemental Table 2).

Attesting to the high quality of the AK58 assembly, 99.97% of the Illumina paired end reads generated in this study could be mapped to the assembly, and the nucleotide accuracy rate of the assembly was 99.9995% based on a homozygous SNP rate of 0.0004968% (Supplemental Table 3). The LTR Assembly Index (LAI) scores of the three subgenomes were all above 10, and the coverage of 15 previously reported BAC

sequences by AK58 assembly reached 99% - 100% (Supplemental Figures 1A and 1B, Supplemental Table 4). CEGMA and BUSCO analysis revealed that 97.2% of the highly conserved eukaryotic coding exons were present in the AK58 assembly (Supplemental Tables 5 and 6).

CENH3 is a functional marker of centromeres in eukaryotes (McKinley and Cheeseman, 2016). Through mapping CENH3 ChIP-Seq reads, we determined the centromere location for all 21 chromosomes (Figure 1, Supplemental Table 7). The average centromere size was 7.0 Mb ranging from 3.0 (7B) to 9.6 Mb (2B) across the 21 chromosomes, with the mean centromere size being 7.5, 7.0 and, 6.7 Mb for A, B, and D subgenomes, respectively. Compared with CS, we observed an increase in centromere size in AK58 chromosomes (Zhao et al., 2023a). Altogether, the above data indicate that AK58 genome is among the well assembled Triticeae genomes reported recently (Supplemental Table 8).

While constructing the 3D map of AK58 genome using 797.6 million pairs of high-confidence Hi-C reads, we noted strong signals along the diagonal of the interaction map, indicative of abundant interactions involving nearby chromosomal regions (Supplemental Figure 1C). Importantly, we revealed that there existed subgenome specific and dominant homologous TEs, which enabled chromosomes of the same subgenome interacted more strongly with each other and thus formed subgenome-specific territories. The 1RS chromosomal arm, introgressed from rye and differing from its wheat counterpart in TE composition, exhibited much less interactions with wheat chromosomes (Jia et al., 2021).

Detailed annotation of genes and TEs in AK58 genome

We annotated 119,448 high-confidence protein coding genes (PCGs) for AK58 genome, with more PCGs in subgenome D (40,665) relative to subgenomes B (38,538) and A (38,115) (Supplemental Table 2). Of the 119,448 PCGs, 117,318 (98.2%) were ordered on the 21 chromosomes (Supplemental Table 9). Consistent with other studies (Athiyannan et al., 2022; Aury et al., 2022), gene density was relatively high towards

the distal regions of the chromosomes where recombination rate was high and TE content was low, with comparatively low gene density found in the pericentromeric region (Figure 1).

Of the 119,448 PCGs annotated in AK58 genome, 63,843 (53.4%) were present as triad homoeologous group (A:B:D configuration of 1:1:1), 21,232 (17.8%) had at least one homoeolog duplicated in one subgenome (A:B:D configuration of 1:1:N, 1:N:1, N:1:1, or other ratio), and 10,192 (8.5%) genes were dyad (loss of homoeolog in one subgenome, resulting in the A:B:D configuration of 1:1:0, 1:0:1, or 0:1:1). The remaining 22,051 (18.5%) genes occurred as singletons (A:B:D configuration of 1:0:0, 0:1:0, or 0:0:1). Lineage-specific gene duplications and pseudogene formation have profoundly shaped the divergence of homoeologous chromosomal loci carrying gluten genes, which collectively determine the end-use quality of wheat grains (Wang et al., 2020). Due to their high complexities, gluten gene loci are usually poorly assembled in previously published wheat genomes (IWGSC et al., 2018; Walkowiak et al., 2020). To aid future wheat evolutionary and grain quality improvement studies, we analyzed the gluten gene loci of AK58 in detail (Supplemental Figure 2, Supplemental Table 10). Two paralogous genes encoding high-molecular-weight glutenin subunits were detected in homoeologous *Glu-A1*, *-B1*, and *-D1* loci, respectively. In the composite locus *Gli-A1*/*Glu-A3*, we found 18 duplicated genes encoding γ-gliadins, ω-gliadins, or low-molecular-weight glutenin subunits (LMW-GSs); on the other hand, 24 duplicated genes coding for γ-gliadins, ω-gliadins, δ-gliadins, or LMW-GSs were detected in homoeologous *Gli-D1*/*Glu-D3* locus. The *Gli-B1*/*Glu-B3* locus was absent in AK58 due to replacement of 1BS by 1RS. Thus, we annotated the *Sec-1* and *Sec-4* loci of the 1RS in AK58; the former carried 21 duplicated genes encoding 40K-γ-secalins whereas the latter had 11 duplicated genes for ω-secalins. As to homoeologous *Gli-A2*, *-B2*, and *-D2* loci on group 6 chromosomes, they carried 35, 18, or 10 duplicated genes specifying α-gliadins. Except for *Glu-B1* and *-D1*, pseudogenization was commonly observed in the other gluten gene loci (Supplemental Table 10).

In AK58, TEs accounted for 85.3% of the genome, of which, retrotransposons and DNA transposons covered 67.09% and 16.56% of the genome, respectively (Supplemental Table 11**)**. Globally, TE content was similar across subgenomes A (85.6%), B (84.6%), and D (82.9%) (Supplemental Figure 3A**)**. TEs were densely distributed in the middle regions of chromosomes where the gene density and recombination rate were low (Figure 1). The accumulative length of TEs in three subgenomes was different for retrotransposons (B > A > D) and DNA transposons (B > D > A). We found that CACTA elements expanded in *Poaceae* species relative to other subfamilies of *Graminaceae*, and accounted for 15.4% of the whole genome and 18.9% of the D subgenome (the highest in the three subgenomes) in AK58, similar to that reported in CS (IWGSC et al., 2018).

Epi-modifications and open chromatins

We investigated whole-genome DNA methylation in single-base resolution in AK58, and found that 116,626 PCGs with methylation in their promoter or gene body regions under normal growth conditions, accounting for 97.6% of the total 119,448 PCGs. Methylation occurred mainly in CG and CHG sites and their levels were positively correlated with TE abundance in promoter regions (Supplemental Figure 3B**)**. In contrast, CHH methylation did not have a clear relationship with TE density, but was preferentially associated with, and likely directed, by the 24-nt small RNAs (Supplemental Figure 3C). We examined histone modification and chromatin accessibility by capturing 19 key histone marks and MNase-digestion accessibility of AK58 genome, providing rich information on histone modifications in common wheat. Overall, approximately 5% of the genome had histone modifications, which occurred in 85,663 (71.7%) of the 119,448 PCGs. Furthermore, chromatin was open in approximately 1.2% of the genome, involving 102,963 (86.2%) of the total 119,448 genes. Our epigenomics data will complement those published previously (Yuan et al., 2020, 2022; Liu et al., 2021; Wang et al., 2021), thus enabling more systematic studies of the roles of epigenetic regulations on wheat trait formation and improvement.

Transcription factors and transcriptional landscape

We annotated 6,355 putative TF genes for AK58 genome, which belonged to 66 families and accounted for 5.32% of the 119,448 PCGs. Notably, the number of TFs in AK58 was evidently more than that in other grass genomes, even polyploid feature was considered (Supplemental Table 12). The number of annotated TF genes in A, B and D subgenome was 2,054, 2241, and 2,060, respectively, with the TF genes in A and D subgenome being more numerous than those annotated for *T. urartu* (1,760) or *Ae. tauschii* (1,892). The top five TF gene families were NAC (549), AP2/ERF (492), C2H2 (491), bHLH (478), and MYB (418), respectively.

To explore the global gene expression patterns of hexaploid wheat, we performed RNA sequencing using the AK58 samples collected from diverse organs, developmental stages, and abiotic stress conditions (Supplemental Table 13). The transcripts for 82,704 genes (69.2% of 119,448 PCGs) were detected and their expression variations among tissues and stress conditions were observed (Supplemental Figure 4). Generally, no obvious subgenome dominance in gene expression was observed. Nevertheless, 10.54% (12,584) of the expressed genes exhibited context-specific expression patterns. We constructed a network based on weighted gene co-expression network analysis and defined 84 co-expression modules (Figure 2A, Supplemental Figure 5), of which 74 may potentially affect wheat growth and/or stress tolerance as they contain one or more rice gene homologs known to function in such processes. A closer inspection revealed that genes from the A, B and D subgenomes were almost equally distributed in each co-expression module (Supplemental Table 14), suggesting that the expression of subgenome orthologs was convergent.

There were probably two major factors to make the expression of subgenome orthologs convergent. One is the TFs that regulated the orthologs in a similar manner across the subgenomes (Figure 2B). Of the 6,355 annotated TF genes, 4,890 were expressed and distributed in the 84 co-expression modules (Supplemental Table 14). The TF genes in each module were co-expressed with the target genes from all three

subgenomes. Co-regulation of TFs and their targets in the three subgenomes plays a key role in yielding the whole genome co-expression network. Another factor might be the genome-wide epigenetic modification that is closely related to gene expression. Although the ancestral diploid progenitors of hexaploid wheat were diverged more than 5 MYA (Marcussen et al., 2014), and the D genome in *Ae. tauschii* and hexaploid wheat for only about 10,000 years (Huang et al., 2002), the epigenetic modifications (mainly histone modifications) were more similar among the three subgenomes of wheat than between the D genome of *Ae. tauschii* and the subgenome D of wheat (Figure 2C), which could contribute to the diploid-like gene expression in hexaploid wheat.

On the other hand, we observed that 41.6% of the AK58 triads expressed in this work displayed expression variations, indicating subfunctionalization of homoeologs according to previous studies (Blanc and Wolfe, 2004; Roulin et al., 2013). This points to the possibility that a specific homoeolog may be preferentially expressed in certain tissues or conditions to assist better perception of developmental cues and/or more efficient environmental adaptation. A notable example was the "Green Revolution" gene *Rht1*. Among its three homoeologs, *Rht-D1* in AK58 encodes a mutant protein, thus leading to constitutive suppression of GA signaling and a desirable dwarf phenotype as reported previously (Peng et al., 1999). Compared with *Rht-A1* and *Rht-B1*, *Rht-D1* was highly expressed in stems (Supplemental Figure 6A), and displayed an obviously higher co-expression pattern with the genes regulating internode architecture, such as the orthologs of *OSH1*, *OSH15*, and *OsSD1* (Supplemental Figure 6B). Another example was the domestication gene *Q* that encodes an AP2-like TF (Zhang et al., 2011). For the three *Q* homoeologs of AK58, *Q-5A* encodes a protein with the V329I mutation, consistent with that reported previously. *Q-5A* was predominantly expressed in developing spikes (Supplemental Figures 6C and 6D), consistent with its function in promoting square spike and higher spikelet density in modern common wheat (Zhang et al., 2011).

SNPs, genetic map, QTLs, functional genes, and mutant library

AK58 reference genome provides a platform for generating and integrating together the SNPs, genetic map, QTLs, and functional genes that are key components for genomics-assisted breeding. We annotated SNP markers in the widely used 55K and 660K chips using AK58 genome information, and mapped 33,124 polymorphic SNP markers onto the AK58/CS F_2 genetic map (Supplemental Figure 7). To add value to this map, we anchored a total of 950 QTLs and 1,227 functionally studied genes published previously to this map; the 1,692 candidate genes under improvement selection revealed by this work, including the 139 HGWAS loci described below, were also integrated (Supplemental Table 15).

We generated 3,031 ethyl methyl sulfone (EMS) mutant M_3 lines for AK58 (Supplemental Figures 8A-8C) and designed an exome capture chip based on the annotated genes of AK58 genome. The exons and introns, as well as the up- and down-stream regions, of genes were captured for 714 EMS lines. A total of 7,193,425 mutations (including 6,033,155 SNPs and 1,160,270 Indels) were precisely identified for 159,184 genes. The EMS-induced SNPs caused 1,342,083 missense, 60,894 stop-gained, and 3,106 start codon-lost mutations in gene coding regions, and a large number of mutations were also found in the untranslated and promoter regions (Supplemental Figure 8D). On average, 6,080 single base mutations were found per line, with eight missense and truncation alleles per gene in this mutant library. The mutation efficiency of AK58 EMS library was similar to those reported previously (Krasileva et al., 2017).

Finally, we built a comprehensive AK58 genome database (available in https://triticeae.henau.edu.cn/aikang58/), and a JBrowse module was developed to view the SNPs, QTLs and multiple epi-modifications. Users may search for nucleotide and deduced protein sequences of their interested genes, and find a wide range of information concerning transcript levels combined with expression modules and co-expression genes, as well as the SNPs, QTLs, and EMS mutations.

Comparisons between AK58 and CS reveal genomic changes in modern

wheat improvement

CS is a well-known landrace in the world, with its genome well assembled and analyzed (IWGSC et al., 2018). The availability of genome sequence of AK58, a modern elite variety extensively cultivated in China, provided us a valuable opportunity for comparing landrace and improved wheat cultivars at a genome-wide level to investigate the genomic changes brought about by modern wheat breeding selection.

The plant architectures were significantly different between AK58 and CS (Figure 3A). AK58 displayed reduced plant height (PH), flag leaf length (FLL) and angle (FLA), heading time (HT), spikelet number per spike (SLN), floret number per spike (FLN), and grain number per spike (GN), but exhibited significant increases in flag leaf width (FLW) and thickness (FLT), awn length (AL), and chlorophyll content (Chl) as well as thousand grain weight (TGW) and harvest index (HI) (Supplemental Table 16). Moreover, the grain quality related traits of AK58, e.g., grain protein and wet gluten contents (GPC and WGC), were also superior over those of CS (Supplemental Table 16). Clearly, the agronomic characteristics of AK58 are typical of those of post green revolution modern elite cultivars, whereas the traits of CS are representative of those of landrace varieties (Hao et al., 2020; Li et al., 2022).

When the pseudochromosomes of AK58 were aligned to those of CS, approximately 86% of the AK58 genome was collinear with 88% of CS genome (Supplemental Table 17). The largest non-syntenic region concerned the short arm of Chr1B because of the 1RS translocation in AK58 (Figure 3B, Supplemental Figure 9). More than 10% genomic difference existed between AK58 and CS, and the B subgenome appeared to be more variable, with PAV genes (genes in present or absent variation) occurred more frequently in the B subgenome than in the A and D subgenomes (Supplemental Table 18).

In total, 40,607,820 SNPs and 5,491,679 Indels existed between AK58 and CS, with variations occurring more frequently in the distal regions than in the peri-centromeric areas of chromosomes (Figure 3C, Supplemental Figure 10). Of these variations,

169,376 SNPs (0.4%) and 12,774 Indels (0.2%) were located in the exons of 36,454 genes, with 83,618 SNPs and 8,259 Indels causing frame-shifting mutations (Supplemental Table 19). Subgenomes B and D had the highest and lowest polymorphisms between the two cultivars based on SNPs and Indels. The SNP frequency between AK58 and CS was 2.8 SNP/kb. Structural variations (SVs) are an important indicator of the evolution and selection that plants have experienced and are thus critical for phenotypic variations (Yuan et al., 2021). We therefore analyzed genotype specific SVs by mutually aligning AK58 and CS genomes. The cumulative lengths of AK58 specific SVs (present in AK58 and absent in CS) and CS private SVs (present in CS and absent in AK58) were 183.3 and 107.8 Mb, respectively, which accounted for 1.3% and 0.8% of AK58 and CS genome, respectively. Within the SVs, 5,857 and 3,080 genes were specifically owned by AK58 and CS, respectively (Supplemental Table 20), accounting for 4.8% and 2.9% of the annotated AK58 and CS PCGs, respectively.

The majority of the SV genes, 75% for AK58 and 66% for CS, were singleton or multiple-copy genes (Supplemental Table 20). GO term enrichment analysis showed both AK58 and CS specific SV genes were enriched in kinase activity and cellular protein modification processes, but more AK58 SV genes were involved in photosystem I and response to wounding. KEGG analysis indicated that AK58 specific SV genes were significantly enriched in oxidative phosphorylation pathway, while CS specific SV genes were mainly involved in plant-pathogen interaction (Supplemental Figure 11).

Based on the transcriptomic data generated using seven organ samples, the expression levels of 8,808 gene pairs were statistically different ($q < 0.05$) between AK58 and CS, and there were more up-regulated genes than down-regulated genes in AK58 relative to CS (Figure 3D, Supplemental Figure 12). These differentially expressed genes were enriched mainly in metabolic process and catalytic activity (Supplemental Figure 13).

Detection of agronomically important homoeologous loci by HGWAS

Following the above analysis, we developed a homoeologous locus-based GWAS (HGWAS) approach to investigate the homoeologous loci controlling 20 important agronomic traits using AK58/CS F_2 population with the phenotypic data collected from two to eight environments. Unlike conventional QTL and GWAS studies that considered the different homoeologs of a homoeologous locus independently during genotyping and association test, HGWAS aimed at establishing marker-trait associations for homoeologous loci, with the different haplotypes of a homoeologous locus defined by combining the SNP markers nearest to each homoeolog (Figure 4A).

Using HGWAS, we detected a total of 393 loci significantly associated with leaf, spike, grain, and plant architecture related traits (Supplemental Table 21), of which 139 were detected in two or more environments (Supplemental Table 22) with significantly different genetic effects for different traits (Figure 4B). Among the 139 HGWAS loci, 123 (88.5%) explained more than 10% of the phenotypic variation each, and were thus regarded as major trait-controlling loci (Supplemental Table 22). The 123 loci were associated with yield components (71), plant architecture (42), heading and maturity times (4), and photosynthesis (6) (Supplemental Table 22). These prominent HGWAS loci distributed on all seven homoeologous groups with some obvious clusters (Supplemental Table 23).

We also performed a conventional GWAS analysis using the same population and same sets of phenotypic data but considered the homoeologs independently using their nearest SNP markers, with the GWAS loci detected compared to those uncovered by HGWAS. A total of 460 loci were detected by the two methods (Supplemental Table 21), including 67 (14.6%) by conventional GWAS only, 219 (47.6%) by HGWAS only, and 174 (37.8%) by both. HGWAS and GWAS analysis detected 85% and 53% of the 460 loci, respectively. Clearly, more than 40% of the loci could not be detected by conventional GWAS analysis (Supplemental Figure 14). For the 174 loci detected by both methods, the percentages of phenotypical variation explained by HGWAS loci were generally higher than those by their corresponding

GWAS loci (Supplemental Table 24), such as heading date (Figure 4C). These data suggest that HGWAS is more powerful than conventional GWAS in discovering agronomically important loci in common wheat.

To validate the high effectiveness of HGWAS, we compared the genetic effects on plot yield of the elite haplotypes of *Vrn3* revealed by conventional GWAS and HGWAS. In common wheat, *Vrn3* (also named as *TaFT-1*), located on group seven chromosomes, exerts pleotropic effects on many important traits including heading date and yield related traits (Yan et al., 2006; Chen et al., 2022). In our single homoeolog-based GWAS analysis of 267 common wheat accessions genotyped using the 660K SNP array, only one of the three homoeologs, i.e., *Vrn3-D1*, was detected to significantly associate with plot yield, with its elite haplotype (Vrn3-7D-hap1) increasing the plot yield by 13.0% relative to the population mean (Table 2). As anticipated, *Vrn3* was found to associate with plot yield in our HGWAS analysis. Among the nine different HHs identified for *Vrn3* in this varietal population, Vrn3-HH1's yield enhancement effect was similar to that of Vrn3-7D-hap1, but importantly we found that two elite HHs, i.e., Vrn3-HH6 and Vrn3-HH7, could both increase the yield by above 30% compared with the population mean (Table 2). These results validate the superiority of HGWAS over single homoeolog-based GWAS in uncovering more elite genetic variants that have much larger genetic effects on agronomic traits, which can contribute directly to the genetic diversities and trait improvement of common wheat.

Identification and analysis of the elite HHs that contributed to modern wheat improvement

For each canonical HGWAS locus with three homoeologs, there were eight homozygous and 19 heterozygous HHs, respectively, in the F_2 population. To analyze the composition of elite HHs for the 123 major HGWAS loci, we designated the parental haplotypes of AK58 and CS as AAA and CCC, respectively, with AAA and CCC indicating the three subgenomic homoeologs of an associated locus all from AK58 or CS. To identify elite HHs, we compared the genetic effects of parental and

505 progeny haplotypes on specific traits. As a control, we used middle parent value
506 (MPV) of the concerned trait, which was the mean of the trait data collected for AK58
507 and CS from multiple environments.

508 To illustrate how this analysis was accomplished, we examined the genetic effects of
509 different HHs in the grain weight associated locus *TGW_G4_16.1_20.0* located in the
510 *Rht1* cluster on PH and TGW (Supplemental Table 25, marked in red). The mean PH
511 of the F_2 plants with the CHA haplotype of *TGW_G4_16.1_20.0*, which carried the
512 CS homoeolog in homozygous state in subgenome A, the AK58 homoeolog in
513 homozygous state in subgenome D, but was heterozygous in subgenome B, was
514 similarly reduced as that of the F_2 individuals with the AAA haplotype (with the three
515 homoeologs of AK58 all in homozygous state). However, CHA increased TGW by
516 6.5% compared to the parental haplotype AAA. Since it is well known that *Rht1* (i.e.,
517 having the AAA haplotype as that in AK58) decreases PH but with a negative effect
518 on TGW (Guan et al., 2018), the CHA haplotype may be useful for mitigating the
519 negative effect of *Rht1* on grain weight while still keeping its PH reduction function
520 in wheat breeding.

521 Using the type of analysis outlined above, we found that, in 54 of the 123 major loci,
522 the AK58 parental AAA haplotypes displayed superior traits compared with the CS
523 parental CCC haplotypes (Supplemental Tables 25 and 26), suggesting that these
524 AAA haplotypes are the products of modern wheat improvement breeding. The 54
525 elite AAA haplotypes carried by AK58 affected plant architecture, yield components,
526 heading time, and photosynthesis, and explained 11% - 40% of the phenotypic
527 variations of the concerned traits. Not surprisingly, these AAA haplotypes concurred
528 with many well characterized important wheat genes (e.g., *Rht1, Rht8, Ppd1*, and
529 *Vrn1*, Supplemental Table 27).

530 For example, in the PH associated locus *PH_G4_15.7_26.2* resided in the *Rht1* cluster,
531 the AAA haplotype (carried by AK58) conferred a 28.8% reduction in PH compared
532 with the CCC haplotype (possessed by CS); in another PH associated locus
533 (*PH_G2_12.2_35.0*) situated in *Rht8* genomic region, the genotype with the AAA

haplotype was about 8 cm shorter relative to that with the CCC haplotype; in the SLN associated locus *SLN_G5_465.8_491.1* located in *Vrn1* genomic region, the F_2 plants with the AAA haplotype had substantially more SLN than those with the CCC haplotype (Supplemental Table 25).

However, the majority of the 54 associated loci possessing elite AAA haplotypes were located in the genomic regions without prior knowledge on agronomically important genes. For example, the GL associated locus *GL_G2_533.3_563.0*, located on the long arm of group 2 chromosomes (Supplemental Figures 15A and 15B), was detected in four environments and explained 11% - 20% of the phenotypic variation of GL (Supplemental Table 21). Notably, *GL_G2_533.3_563.0* overlapped with the grain weight associated locus *TGW_G2_543.6_565.8* (Supplemental Table 25), consistent with the contribution of GL to TGW. For both loci, the elite AAA haplotype was superior over both the CCC haplotype and MPV as AAA gave rise to substantially higher GL and TGW values (Supplemental Figure 15C, Supplemental Table 25). Within this genomic region on Chromosome 2D of AK58, 28 genes (*TraesAK58CH2D473400 - 476100*) were annotated, 11 of which were found expressed in seven tissues with identical patterns between AK58 and CS (Supplemental Figure 15D). The gene *TraesAK58CH2D475100*, predicted to encode an uncharacterized protein with a SMR-domain, showed high transcriptional levels in the spikes and anthers and the grains at early development stages (Supplemental Figures 15D-15F). To seek for genetic evidence for the function of *TraesAK58CH2D475100* in GL and TGW control, we made use of the EMS mutant library of AK58 (see above). Five independent homozygous EMS mutants for *TraesAK58CH2D475100* were identified in the mutant library, and they all produced significantly smaller grains compared with WT AK58 (Supplemental Figure 16). This analysis not only reveals *TraesAK58CH2D475100* as a valuable candidate gene for controlling wheat GL and TGW, but also demonstrates the high utility of AK58 genome resources generated by this work.

Potential superior HH haplotypes for further wheat improvement

Among the 123 major HGWAS loci, 83 carried the HHs superior over the parental AAA haplotypes (carried by AK58) in their genetic effects on the agronomic traits examined in this work (Supplemental Table 26). Of the superior HHs in the 83 loci, 22 were carried by the CS parent (i.e., CCC haplotypes), whereas 61 were newly formed in the F_2 progenies by reassortment of CS and AK58 homoeologs. That is, the 61 HHs were called as BPV ones for their conferring superior traits to the better parent value.

The length, width, and thickness of flag leaves are the architecture traits highly important for wheat yield (Zhao et al., 2018; Tu et al., 2021). As illustrated by AK58 (Figure 3, Supplemental Table 16), the flag leaves of modern variety were shorter, wider, and thicker than those of landraces. In the FLL associated locus *FLL_G4_481.0_485.3*, the superior haplotype AAC (with the A and B homoeologs from AK58 and the D homoeolog from CS) shortened FLL by 3.82 cm (18%) and 2.63 cm (14%) compared with the parental haplotypes CCC and AAA, respectively (Supplemental Table 25). Interestingly, *FLL_G4_481.0_485.3* co-located with the GN associated locus *GN_G4_474.8_485.3*, and the AAC haplotype increased GN by 5.6 (10%) and 5.9 (11%) in comparison with CCC and AAA, respectively (Supplemental Table 25). Therefore, it is worthy to further explore the value of the AAC haplotype of *FLL_G4_481.0_485.3* (*GN_G4_474.8_485.3*) in improving wheat plant architecture and grain yield. *PH_G4_299.3_319.2*, located on the long arm of group 4 chromosomes, was a major locus associated with PH, and its ACA haplotype (with the A and D homoeologs from AK58 and the B homoeolog from CS) shortened PH by 17.6 cm (16.4%) compared with the parental haplotype CCC. More importantly, it had no negative effect on TGW as *Rht1* did (Supplemental Table 25). Therefore, the ACA haplotype of *PH_G4_299.3_319.2* may replace *Rht1* in future wheat breeding.

Vrn1 plays a pivotal role in flowering time control in common wheat (Chen and Dubcovsky, 2012), and allelic variations of *Vrn1* have been reported to cause differences in flowering time (Strejčková et al., 2021). Remarkably, we detected nine HGWAS loci in the genomic region of *Vrn1*, which were significantly associated with

plant architecture, yield components, heading and maturity dates, and photosynthesis (Table 3, Supplemental Table 25). Among the elite HHs of the nine HGWAS loci overlapping with *Vrn1*, the CAA haplotype of *GN_G5_476.7_492.6* had five more grains per spike than the better parent haplotype CCC and seven more grains per spike than MPV; the AAC haplotype of *Ht_G5_473.4_490.5* headed five days earlier than MPV (Table 3, Supplemental Table 25).

The above data prompted us to further examine *Vrn1* HHs using more diversified common wheat germplasm. We thus determined *Vrn1* HHs in 77 landraces and 337 improved varieties, which unveiled a total of 50 *Vrn1* HHs (Supplemental Table 28). The dominant *Vrn1* locus haplotypes were HH13 and HH5 in both landraces and improved varieties, but HH20 and HH35 were also abundant in the examined landraces. The most favorable was HH33, which reduced PH by 19.0% but increased TGW by 4.4% and grain yield by 16.0% compared with the dominant haplotype HH5 in three-year field trials. HH16 carried by AK58 also reduced PH by 12.9% and promoted TGW by 5.4% and grain yield by 4.1% relative to HH5. Thus, HH16 was not as effective as HH33 in promoting the grain yield of common wheat. Notably, HH33 was a rare HH of *Vrn1*, as it was detected in only 3.3% of the varieties analyzed here (Supplemental Table 28).

Discussion

In this work, we built the genome database of AK58, an elite winter type variety developed by intensive selection (Wang et al., 2018; Jia et al., 2021). The assembled genome size of AK58 (14.75 Gb) was comparable to that reported for other varieties, e.g., 14.77 Gb for Kenong 9204 (Shi et al., 2022) and 14.96 Gb for SY Mattis (Walkowiak et al., 2020). Although the contig and scaffold parameters of AK58 assembly were lower than those of the common wheat genome assemblies reported very recently (Athiyannan et al., 2022; Aury et al., 2022; Kale et al., 2022), the quality of AK58 genome assembly was similar to that of Kenong 9204 released in 2022 (Shi et al., 2022). Importantly, the genome database of AK58 is more

wide-ranging than that reported for previously sequenced common wheat varieties. The 3D chromatin architecture of Aikang 58 was already proved to be useful for revealing homology-mediated inter-chromosomal interactions in hexaploid wheat (Jia et al., 2021). The data generated in this work further demonstrated the high utility of AK58 genome resources.

Owing to its hexaploid nature, large genome size, and high percentage of TEs, the epigenetic studies of common wheat have lagged behind those of model plants (Arabidopsis and rice) and many crop species (Song et al., 2017; Zhao et al., 2020; Jiang et al., 2021; Samantara et al., 2021). Nevertheless, common wheat is a unique and powerful model for studying the important roles of epigenetic regulations in crop evolution and improvement (Yuan et al., 2020, 2022). The rich and systematic epigenomics data and the HGWAS loci reported here for AK58, plus the epigenetic resources generated previously (Yuan et al., 2020, 2022; Liu et al., 2021; Wang et al., 2021), will provide a solid basis and practical clues for fast tracking the research on the functions of epigenetic regulations on trait formation and enhancement in common wheat in the future.

Through intensive selection breeding, semi-dwarf modern cultivars, resistant to lodging but requiring more nitrogen fertilizers to achieve high yield level, become prevalent in global wheat production (Hedden, 2003; Wu et al., 2020). Recent studies indicate that synergistic selection of the genes with multiple functions and pleiotropic effects plays an important part in shaping the performance of modern wheat cultivars (Hao et al., 2020; Pang et al., 2020; Li et al., 2022), and that introduction of alien genes from wheat relatives aided the resilience of common wheat under adverse environmental conditions (Qi et al., 2007; Mirzaghaderi and Mason, 2019; Walkowiak et al., 2020; Zhang et al., 2023). Nevertheless, the effects of modern breeding selection are very complex; a complete understanding of the genetic, molecular, and physiological changes involved is still far away (Shi and Lai, 2015).

Comparison of AK58 with CS in this work indicates that cultivar specific SVs may contribute to AK58's superior performance over CS. SVs rendered AK58 to possess a much higher number of private genes (5,758) than CS does (3,080), with many of the AK58 specific genes involved in photosystem I and oxidative phosphorylation according to GO or KEGG analysis. Because both processes are involved in producing ATP through photosynthesis in chloroplasts or respiration in mitochondria, AK58 may have a higher cellular content of ATP than CS. As ATP is the most important source of energy in cells, an enhanced supply of ATP may enable AK58 to grow, and to defend against environmental stresses, more robustly, and thus achieving higher and more stable yield levels under different growth conditions. Obviously, our comparative analysis of AK58 and CS genomes provides valuable clues for systematically dissecting the genetic, molecular and physiological basis of modern breeding on wheat improvement.

In common wheat, there is so far only one report in the literature that has identified and compared the genetic effects of different HHs on agronomic traits. In their study, Dong et al. (2010) compared the genetic effects of eight HHs formed by reassortment of parental *Glu-A3*, *-B3* and *-D3* homoeologs on the gluten quality parameter Zeleny sedimentation value (ZSV), and identified a superior progeny HH (with *Glu-A3* and *-D3* from one parent and *Glu-B3* from another parent) using PCR markers, whose ZSV was 21.96% higher than MPV and 5.97% higher than BPV. This illustrates the possibility, importance and high potential of obtaining elite HHs for improving agronomic traits. Herein, we proved that homoeologous locus-based HGWAS is substantially more effective than single homoeolog-based GWAS in discovering the chromosome loci and their elite HHs controlling important agronomic traits (Table 2, Supplemental Figure 14, and Supplemental Tables 21 and 24). Of the 123 major HGWAS loci detected by us, many acted pleiotropically on two or more important traits (e.g., PH and TGW, GL and TGW, or FLL and GN). This is in accordance with the finding that modern wheat breeding has synergistically selected multiple key genes with pleiotropic effects (Li et al., 2022). Through analyzing the genetic effects

of the 123 major HGWAS loci using MPV as control, we deduce that the 54 loci, whose elite HHs were AK58 parental homoeolog sets, are very likely the products of modern intensive selection breeding. This is consistent with the finding that many of the 54 loci were located in the genomic regions harboring well known wheat improvement genes (e.g., *Rht1*, *Rht8*, *Ppd1*, and *Vrn1*) (Supplemental Table 27). This proposition is also supported by the identification of *TraesAK58CH2D475100* as a candidate gene for the two overlapping loci associated with GL (*GL_G2_533.3_563.0*) and GW (*TGW_G2_543.6_565.8*), respectively (Supplemental Figures 15 and 16).

In contrast to the 54 loci discussed above, the 83 HGWAS loci, whose elite HHs conferred higher genetic effects than BPVs, may be valuable for further improvement of common wheat. In 21 such loci, the elite haplotypes were from CS. This may not be surprisingly, as landrace cultivars have often been found to carry elite alleles for better environmental adaptability, more potent defense responses to stresses, or superior quality parameters in crop genetic studies (Liu et al., 2019; Rufo et al., 2019). Remarkably, in 62 loci (~75% of the 83 loci), the elite HHs were newly formed by reassortment of CS and AK58 homoeologs, indicating that the likelihood of obtaining favorable HHs with high breeding values is quite large. Of particular interest is the detection of nine HGWAS loci in the genomic region of *Vrn1*, with many elite HHs conferring superior agronomic traits (Supplemental Tables 25 and 27). Consistently, elite *Vrn1* HHs were also discovered in wheat landrace and improved cultivars, with the rare haplotype HH33 reducing PH by 19.0% and simultaneously increasing TGW by 4.4% and grain yield by 16.0% compared with the dominant haplotype HH5 in multi-year field trials (Supplemental Table 28). Thus, the elite *Vrn1* HHs discovered in this work, especially HH33, may help to revolutionize wheat yield improvement if introduced into appropriate genetic background through genomics-assisted breeding in the future.

Our work showed that HGWAS was more powerful than conventional GWAS analysis in terms of the number of positive loci identified and the PVE% explained by

the associated loci. This is understandable as HGWAS treated the three subgenomic homoeologs as one and thus has a higher probability of detecting the synergistic function of the three homoeologs. Another observation was that the HGWAS loci associated with different traits tended to form clusters. This may be caused by the neofunctionalization of one or more of the three subgenomic homoeologs. With these considerations, the HGWAS loci may aid further investigations of the conserved and diverged functions of wheat homoeologs as well as their additive and nonadditive interactions in agronomic trait control in future research. Finally, since the genetic effects of a HGWAS locus may reflect the combined function of three subgenomic homoeologs, the HGWAS approach and the loci revealed using it might also help to stimulate further and deeper studies of the genetic basis of polyploid heterosis, a phenomenon often exhibited by polyploids when compared with their progenitors with lower ploidy levels (Abel et al., 2005; Chen et al., 2010; Bansal et al., 2012).

Previous studies suggest that the genetic diversity of hexaploid wheat is very poor, and this problem has been regarded as a bottleneck limiting the progress of wheat improvement (Mirzaghaderi and Mason, 2019; Scott et al., 2021). However, using the HGWAS approach, we demonstrate that HH variations are very rich in hexaploid wheat, with the probability of identifying elite HHs being fairly high. Hence, discovery and application of elite HHs may lead to breakthroughs in wheat breeding programs in the future.

In summary, our work has generated a valuable genome database for an elite common wheat variety, which enriches wheat genomic resources and may contribute positively to worldwide wheat genomics, germplasm enhancement, and breeding studies. The insights obtained using AK58 genomic data highlight the potential benefits of HGWAS and the elite HHs mined by HGWAS, whose further testing and efficient exploitation will likely enhance the genetic diversity and accelerate genomics-assisted breeding in common wheat and other polyploid crops.

Methods

Plant materials and growth conditions

AK58 was provided by its breeder Professor Zhengang Ru. CS was the line sequenced previously (IWGSC et al., 2018). Wheat plants were grown under greenhouse conditions with day and night temperatures of 25 °C and 20 °C, respectively, and a photoperiod 16 h light / 8 h dark. The AK58 × CS F_2 population was prepared using AK58 as the female parent. AK58 and CS, as well as their F_2 individuals and derivative $F_{2:3}$ families, were cultivated in multiple environments for phenotypic data collection as reported previously (Zhang et al., 2013; Zhao et al., 2018; Tu et al., 2021, detailed below). The processing quality-related parameters of AK58 and CS were scored as reported by Zhang et al. (2018).

Genome sequencing

AK58 genomic DNA was used to construct multiple types of libraries, including short insert size (450 bp) libraries, mate-paired (2 kb, 5 kb, 8 kb, 20 kb and 40 kb) libraries and PacBio SMRT Cell libraries. For the 450 bp short inserts, PCR free libraries were constructed according to the manufacturer's instructions and sequenced on an Illumina HisSeq2500 instrument with 250 bp per end. The libraries with different fragment sizes ranging from 2 to 40 kb were constructed and sequenced on the Illumina X Ten platform. PacBio SMRT Cell libraries were sequenced with a PacBio RS II instrument.

Genome assembly and evaluation

AK58 genome assembly was accomplished using the software package DeNovoMAGIC2 (NRGene, Nes Ziona, Israel), which is highly efficient in assembling the genomes rich in repetitive elements (https://www.nrgene.com/de-novo-magic/). Sequencing data from PCR-Free library and the Nextera mate-paired libraries were used for DeNovoMAGIC2 assembly. PCR duplicates, an Illumina adaptor (AGATCGGAAGAGC), and Nextera linkers (for mate-paired libraries) were removed from raw sequencing data. Overlapping reads

from the PE 450 bp 2 × 250 bp libraries were then merged with a minimal required overlap of 10 bp to create the stitched reads. The first step of the DeNovoMAGIC2 assembly algorithm consisted of building a De Bruijn graph (kmer = 191 bp) of contigs from the overlapping PE reads. Next, PE reads were used to find reliable paths in the graph between contigs for repeat resolving and contig extension. Later, contigs were linked into scaffolds with PE and MP information, estimating gaps between the contigs according to the distance of PE and MP links. The final fill gap step used PE and MP links, as well as De Bruijn graph information, to detect a unique path connecting the gap edges. Mate-paired data (20 kb, 40 kb) were mapped to the basic assembly using bowtie (http://bowtie-bio.sourceforge.net/index.shtml), and only unique mapping reads were used for further scaffolding, which was performed by SSPACE (https://www.baseclear.com/ genomics/bioinformatics/basetools/SSPACE, V3.0). PBJelly (http://www.winsite.com/ Home-Education/Science/PBJelly/) was used to fill gaps using approximately 10× of PacBio SMRT sequencing data. The high-density genetic map of AK58 × CS was used to anchor the scaffolds to chromosomes using BLAST program. The completeness of gene regions of the assembly was evaluated using both CEGMA (Core Eukaryotic Gene Mapping Approach, http://korflab.ucdavis.edu/datasets/cegma/) and BUSCO (Benchmarking Universal Single-Copy Orthologs, http://busco.ezlab.org/). LAI scores were computed for A, B and D subgenomes, respectively, to assess the quality of the assembly of intergenic regions (Ou et al., 2018). To examine the accuracy of AK58 assembly, the raw Illumina reads were aligned to AK58 genome using BWA software. Then alignments were sorted using SAMtools, and the variants were called using GATK HaplotypeCaller module. The SNPs were filtered by use of VCFtools. Homozygous SNPs were used to calculate nucleotide base accuracy rate of the assembly.

Protein-coding gene prediction

Protein-coding region identification and gene prediction were conducted using a combination of homology-based prediction, *de novo* prediction, and transcriptome-based prediction methods. Protein sequences from eight grass genomes

(*Brachypodium distachyon*, *Sorghum bicolor*, *Oryza sativa*, *Zea mays*, *Hordeum vulgare*, *T. urartu*, *Setaria italic*, *Panicum virgatum*) were downloaded from Ensemble (release-33). Protein sequences of three additional Triticeae species were downloaded from the websites https://www.ncbi.nlm.nih.gov/nuccore/AOCO02000000 (for *Ae. tauschii*), http://wewseq.wixsite.com/consortium/wild-emmer-wheat (for *T. turgidum* ssp. *diccocoides*), or https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_ Annotations/v1.0/ (for CS, IWGSCv1.0). The protein sequences from the above eleven genomes were aligned to AK58 genome assembly using TblastN with an E-value cutoff of 1e-5. The BLAST hits were conjoined using Solar software. GeneWise was used to predict the exact gene structure of the corresponding genomic regions for each BLAST hit. Homology predictions were split into two sets, which included a high-confidence homology set (HCH-set, with significant identities to the genes annotated in CS) and a low confidence homology set (LCH-set, all except for the HCH-set). A collection of wheat full-length cDNAs (16,807 in total) were directly mapped to the AK58 genome and assembled by PASA. Gene models created by PASA were denoted as the PASA-FLC-set (PASA full length cDNA set), this gene set was used to train the *ab initio* gene prediction programs. Five *ab initio* gene prediction programs, i.e., Augustus (version 2.5.5), Genscan (version 1.0), GlimmerHMM (version 3.0.1), Geneid, and SNAP, were used to predict coding regions in the repeat-masked genome. RNA-seq data were mapped to the assembly using ToHILSt (version 2.0.8). Cufflinks (version 2.1.1) was then used to assemble the transcripts into gene models (Cufflinks-set). In addition, 56.51 Gb of RNA-seq data from seven different organs (leaf, root, node, internode, sheath, young spike, and developing grain) were assembled by Trinity, creating several sets of expressed sequence tags (ESTs). These ESTs were also mapped to the AK58 assembly and gene models were predicted using PASA. This gene set was denoted as PASA-T-set (PASA Trinity set). Gene model evidence from the HCH-set, LCH-set, PASA-FLC-set, Cufflinks-set, PASA-T-set and *ab initio* programs were combined by EvidenceModeler (EVM) into a non-redundant set of gene structures. Weights for

each type of evidence were set as follows: HCH-set > PASA-FLC-set > PASA-T-set > Cufflinks-set > LCH-set > Augustus > GeneID = SNAP = GlimmerHMM = Genscan. Gene model output by EVM with low confidence scores was firstly filtered by two criteria: (1) coding region lengths of 150 bp and (2) supported only by *ab initio* methods and with FPKM < 1. Using a similar approach as described in the genome sequencing of *Gossypium raimondii* (Wang et al., 2012), we further filtered gene models based on Cscore and peptide coverage, followed by overlapping CDS with TEs. Only the transcripts with a Cscore ≥ 0.5 and peptide coverage ≥ 0.5 were retained. For the gene models with more than 20% of their CDS overlapping with TEs, we required that their Cscore values must be ≥ 0.8 and that peptide coverage must be ≥ 80%. Finally, we also filtered out those models for which more than 30% of the peptides were annotated as Pfam or Interprot TE domains. Finally, 119,448 high-confidence PCGs were annotated.

Transcriptome sequencing and analysis

To facilitate gene annotation and investigation of biological questions using gene expression information, 42 transcriptomic datasets were generated for AK58 by performing Illumina RNA sequencing of the leaf samples collected from normal or diverse abiotic conditions, the stem, root, and spike samples of normally-grown plants, and the developing grain samples harvested at 4, 10, 15, 20 d after anthesis (Supplemental Table 13). RNA-seq data were mapped to the genome assembly using ToHILSt (version 2.0.8). Only the aligned reads located within 600 bp of each other were defined as concordantly mapped pairs, which were used in the downstream quantification analysis. The minimum and maximum intron length was set to 5 bp and 50,000 bp, respectively. All other parameters were set to the default values. The software cufflinks30 (version 2.1.1) (http:// cufflinks.cbcb.umd.edu/) was used to estimate the expression level for each gene based on the reads uniquely mapped to the genome assembly. An expressed gene was defined if its RPKM value was ≥ 1. Those with an RPKM value < 1 were considered as non-expressed genes. The expressed PCGs were used to build co-expression network using the WGCNA R package

following the study by Yang et al. (2021). Co-expression network was visualized with Cytoscape (Version 3.5.1).

TF analysis

The iTAK program was used to annotate the TF genes of AK58 based on homology search against the known plant TF database integrated in the program, with the search results classified into different TF families. Comparison of AK58 TF genes with those of other grass species were carried out as reported by Zheng et al. (2016).

TE annotation

The complement of AK58 TEs was annotated through homology-based prediction method. A TE library containing 3,050 complete TE sequences (ClariTeRep) was downloaded (https://github.com/jdaron/CLARI-TE). This library was constructed from two curated Triticeae TE libraries: TREP and an additional set of Tes manually annotated in a pilot study of chromosome 3B. This combined library was searched against the AK58 genome using RepeatMasker (https://www.repeatmasker.org/).

Genomic comparison between AK58 and CS

Each pseudochromosome of AK58 genome was aligned to the corresponding chromosome of CS using the software MuMmer (Kurtz et al., 2004). Approximately 12.339 Gb (86.01% of the 14.346 Gb) from AK58 genome could be aligned to 12.311 Gb (87.53% of the 14.066 Gb) of the CS genome, with the average identity of the aligned regions reaching 99.66%.

Detection of HGWAS loci

A total of 1,045 F_2 progenies, derived by crossing AK58 with CS, were evaluated in two field environments during 2018 - 2019, i.e., Xinxiang, Henan province (E113°48′28″, N35°09′34″, 374 F_2 plants phenotyped) and Beijing (E116°20′04″, N39°58′02″, 222 F_2 plants phenotyped), and in two greenhouse experiments in Beijing (one in 2018 involving 238 F_2 plants and another in 2019 with 211 F_2 plants). In addition, 717 $F_{2:3}$ lines were also phenotyped in three field environments during 2019 - 2020, including 200 lines sown on October 15 and 259 lines sown on December 1, 2019 in Xinxiang, Henan Province, as well as 258 lines sown in 2019 in

Changping, Beijing (E116°14′49″, N40°10′48″). A total of 29 traits were recorded as list in Supplemental Table 21. Field management was performed according to the common practices for wheat production. HT was recorded as the days from sowing to heading. At heading stage, FLL, FLW, FLT, LA and Chl were measured using 10 randomly selected flag leaves (Zhao et al., 2018; Tu et al., 2021). At physiological maturity, PH and yield related traits were scored as described by Zhang et al. (2013).

The 1,045 F_2 plants and 717 $F_{2:3}$ lines were genotyped using the 55K SNP array by China Golden Marker (Beijing, China) (Zhai et al., 2021), with a total of 53,063 SNPs identified. Quality control of markers was performed to exclude those with high missing rate (> 50%) and low MAF (< 5%). Using the resultant high-quality SNPs and based on the genome annotation information of AK58 and CS, we identified 17,783 triad gene sets whose homoeologs were polymorphic between AK58 and CS based on the SNP nearest to each homoeolog. These polymorphic homoeologous loci were then used in HGWAS and conventional GWAS analyses using the phenotypic data from each environment. In HGWAS, different genotypes were distinguished to the level of homoeologous loci, whereas in GWAS each homoeolog was genotyped independently. Genotype-phenotype association was tested using the mixed linear model, with population structure and kinship coefficients calculated by the TASSEL3.0 software (Yu et al., 2006). Only the associations with a -log10 (p-value) ≥ 3.0 were selected for further uses. To identify elite haplotypes, the genetic effects of different HHs of the concerned HGWAS locus were compared to MPV or BPV, with statistical analyses accomplished using either Student t-test or LSD multiple comparison test installed in SPSS for windows 13.0.

Investigation of *Vrn1* and *Vrn3* HHs in wheat varietal populations

For investigating *Vrn1* HHs, a total of 414 accessions (including 77 landraces and 337 improved varieties), which had been phenotypically assessed in multiple environments by Gao et al. (2017), were genotyped using the 660K SNP array as described previously (Sun et al., 2020). The resulting SNP data were used to distinguish different HHs of *Vrn1* as described above. The genetic effects of different

HHs on agronomic traits were then computed using the phenotypic data collected previously (Gao et al., 2017). The *Vrn3* HHs and their genetic effects on agronomic traits were investigated similarly, except that the number of varietal accessions used was 267, which were part of the 414 accessions described above.

Funding

This project was supported by the Collaborative Innovation Center for Henan Grain Crops, the Ministry of Science and Technology of China (2021YFF1000200), the National Natural Science Foundation of China (Major Program, 31991213), the Central Public-interest Scientific Institution Basal Research Fund (Y2021YJ01), the Major Public Welfare Projects of Henan Province (201300110800), the Key Research and Development Program of China (2016YFD0100102), CAAS Agricultural Science and Technology Innovation Program (CAAS-ZDRW202002), the seed innovation program of the Ministry of Agriculture and Rural Affairs of China, and Henan Provincial R&D Projects of Inter-regional Cooperation for Local Scientific and Technological Development Guided by Central Government (YDZX20214100004191).

Author contributions

L.G., D.W., Z.R. and X.K. initiated the project and designed the study. J.J., L.G. and D.Y. performed HGWAS and comparative genomic analyses. G.Z. analyzed TF gene and maintained sequence data. D.L. took part in HGWAS and GWAS analysis. G.Z. and K.W. executed genome sequencing and assembly. K.W. performed gene annotation and comparative genomic analyses. C.K. analyzed histone modification and chromatin accessibility. P.D. and L.W. performed transcriptome data analysis. X.Y. and Y.J. performed gene duplication detection and analysis. X.Z. analyzed the centromeres. Z.L. performed data analysis and established a comprehensive genomic database. S.X. and K.C. performed *Vrn1* gene analysis. D.C., C.D., T.L., K.Z. and F.C. maintained laboratory supplies and greenhouse conditions and performed field trials.

G.L. analyzed the structure of gluten loci and gluten genes. Y.Z. analyzed and constructed 3D map. L.Z., X.L. and X.K. developed the F_2 genetic population and assisted phenotypic data collection. Z.R. developed and provided the initial seeds of wheat variety AK58. J.J., L.G., D.W., Z.W. and G.Z. wrote the paper. All authors read and approved the manuscript.

Acknowledgements

No conflict of interest is declared.

Data availability

The genome sequence data reported in this paper have been deposited in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences, under accession number GWHANRF00000000 publicly accessible at https://bigd.big.ac.cn/gwh. The raw sequence data of the genome assembly and transcriptome reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA013077) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The raw data of epigenetics datasets including of ChIP-seq and MNase-seq used in this study can be available in the National Genomics Data Center (NGDC, https://bigd.big.ac.cn) under project accession number PRJCA012697. Other data and materials supporting the findings of this study are available from the corresponding authors upon request.

References

Abel, S., Mollers, C., and Becker, H.C. (2005). Development of synthetic *Brassica napus* lines for the analysis of "fixed heterosis" in allopolyploid plants. Euphytica **146:** 157-163..

Akpinar, B.A., Leroy, P., Watson-Haigh, N., Baumann, U., Barbe, V., and Budak, H. (2022). The complete genome sequence of elite bread wheat cultivar, "Sonmez". F1000Research **11:** 614-614.

Athiyannan, N., Abrouk, M., Boshoff, W.H.P., Cauet, S., Rodde, N., Kudrna, D., Mohammed, N., Bettgenhaeuser, J., Botha, K.S., Derman, S.S., et al. (2022). Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning. Nat. Genet. **54:** 227-231.

Aury, J.-M., Engelen, S., Istace, B., Monat, C., Lasserre-Zuber, P., Belser, C., Cruaud, C., Rimbert, H., Leroy, P., Arribat, S., et al. (2022). Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. Gigascience **11:** giac034.

Bansal, P., Banga, S., and Banga, S.S. (2012). Heterosis as investigated in terms of polyploidy and genetic diversity using designed *Brassica juncea* amphiploid and its progenitor diploid species. PLoS One **7:** e29607.

Blanc, G., and Wolfe, K.H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679-1691.

Chen, A., and Dubcovsky, J. (2012). Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. PLoS Genet. **8:** e1003134.

Chen, Z., Ke, W., He, F., Chai, L., Cheng, X., Xu, H., Wang, X., Du, D., Zhao, Y., Chen, X., et al. (2022). A single nucleotide deletion in the third exon of *FT-D1* increases the spikelet number and delays heading date in wheat (*Triticum aestivum* L.). Plant Biotechnol. J. **20:** 920-933.

Chen, Z.J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci. **15:** 57-71.

Dong, L., Zhang, X., Liu, D., Fan, H., Sun, J., Zhang, Z., Qin, H., Li, B., Hao, S., Li, Z., et al. (2010). New insights into the organization, recombination, expression and functional mechanism of low molecular weight glutenin subunit genes in bread wheat. PLoS One **5:** e13548.

Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. Science **316:** 1862-1866.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. Cell **184**: 1621-1635.

Gao, L., Zhao, G., Huang, D., and Jia, J. (2017). Candidate loci involved in domestication and improvement detected by a published 90K wheat SNP array. Sci. Rep. **7:** 44530.

Guan, P., Lu, L., Jia, L., Kabir, M.R., Zhang, J., Lan, T., Zhao, Y., Xin, M., Hu, Z., Yao, Y., et al. (2018). Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable loci for yield-related traits across different environments in wheat (*Triticum aestivum* L.). Front. Plant Sci. **9:** 529.

Gupta, P.K., Kulwal, P.L., and Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. Adv. Genet. **104:** 75-154.

Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y., Zheng, J., Liu, H., Bi, Z., Xu, F., et al. (2020). Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. Mol. Plant **13:** 1733-1751.

He, Z., Xia, X., Peng, S., and Lumpkin, T.A. (2014). Meeting demands for increased cereal production in China. J Cereal Sci. **59:** 235-244.

Hedden, P. (2003). The genes of the Green Revolution. Trends Genet. **19:** 5-9.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., and Gornicki, P. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proc. Natl. Acad. Sci. USA **99:** 8133-8138.

IWGSC., Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Choulet, F., Distelfeld, A., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science **361:** eaar7191.

Jackson, S., and Chen, Z.J. (2010). Genomic and expression plasticity of polyploidy. Curr. Opin. Plant Biol. **13:** 153-159.

Jia, J., Xie, Y., Cheng, J., Kong, C., Wang, M., Gao, L., Zhao, F., Guo, J., Wang, K., Li, G., et al. (2021). Homology-mediated interchromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. Genome Biol. **22:** 26.

Jiang, X., Song, Q., Ye, W., and Chen, Z.J. (2021). Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. Nat. Ecol. Evol. **5:** 1382-1393.

Kale, S.M., Schulthess, A.W., Padmarasu, S., Boeven, P.H.G., Schacht, J., Himmelbach, A., Steuernagel, B., Wulff, B.B.H., Reif, J.C., Stein, N., et al. (2022). A catalogue of resistance gene homologs and a chromosome-scale reference sequence support resistance gene mapping in winter wheat. Plant Biotechnol. J. **20:** 1730-1742.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X., Borrill, P., et al. (2017). Uncovering hidden variation in polyploid wheat. Proc. Natl. Acad. Sci. USA **114:** E913-E921.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. **5:** R12.

Levy, A.A., and Feldman, M. (2022). Evolution and origin of bread wheat. Plant Cell **34**: 2549-2567.

Li, A., Hao, C., Wang, Z., Geng, S., Jia, M., Wang, F., Han, X., Kong, X., Yin, L., Tao, S., et al. (2022). Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. Mol. Plant **15:** 504-519.

Liu, J., Rasheed, A., He, Z., Imtiaz, M., Arif, A., Mahmood, T., Ghafoor, A., Siddiqui, S.U., Ilyas, M.K., Wen, W., et al. (2019). Genome-wide variation patterns between landraces and cultivars uncover divergent selection during modern wheat breeding. Theor. Appl. Genet. **132:** 2509-2523.

Liu, Y., Yuan, J., Jia, G., Ye, W., Chen, Z., and Song, Q. (2021). Histone H3K27 dimethylation landscapes contribute to genome stability and genetic recombination during wheat polyploidization. Plant J. **105:** 678-690.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155.

Ma, X.F., and Gustafson, J.P. (2005). Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenet. Genome Res. **109:** 236-249.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B., Steuernagel, B., Mayer, K., and Olsen, O.A. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. Science **345:** 1250092.

May, D., Paldi, K., and Altpeter, F. (2023). Targeted mutagenesis with sequence-specific

nucleases for accelerated improvement of polyploid crops: Progress, challenges, and prospects. Plant Genome**:** e20298.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat. Rev. Mol. Cell Biol. **17:** 16-29.

Michael, T.P., and VanBuren, R. (2015). Progress, challenges and the future of crop genomes. Curr. Opin. Plant Biol. **24:** 71-81.

Mirzaghaderi, G., and Mason, A.S. (2019). Broadening the bread wheat D genome. Theor. Appl. Genet. **132:** 1295-1307.

Moghe, G.D., and Shiu, S.H. (2014). The causes and molecular consequences of polyploidy in flowering plants. Ann. N. Y. Acad. Sci. **1320:** 16-34.

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46:** e126.

Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., He, F., Li, L., Wang, L., Yuan, X., et al. (2020). High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol. Plant **13:** 1311-1327.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. Nature **400:** 256-261.

Przewieslik-Allen, A.M., Wilkinson, P.A., Burridge, A.J., Winfield, M.O., Dai, X., Beaumont, M., King, J., Yang, C.Y., Griffiths, S., Wingen, L.U., et al. (2021). The role of gene flow and chromosomal instability in shaping the bread wheat genome. Nat. Plants 7**:** 172-183.

Purugganan, M.D., and Jackson, S.A. (2021). Advancing crop genomics from lab to field. Nat. Genet. **53:** 595-601.

Qi, L., Friebe, B., Zhang, P., and Gill, B.S. (2007). Homoeologous recombination, chromosome engineering and crop improvement. Chromosome Res. **15:** 3-19.

Roulin, A., Auer, P.L., Libault, M., Schlueter, J., Farmer, A., May, G., Stacey, G., Doerge, R.W., and Jackson, S.A. (2013). The fate of duplicated genes in a polyploid plant genome. Plant J. **73:** 143-153.

Rufo, R., Alvaro, F., Royo, C., and Miguel Soriano, J. (2019). From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. PLoS One **14:** e0219867.

Salman-Minkov, A., Sabath, N., and Mayrose, I. (2016). Whole-genome duplication as a key factor in crop domestication. Nat. Plants **2**: 16115.

Samantara, K., Shiv, A., de Sousa, L.L., Sandhu, K.S., Priyadarshini, P., and Mohapatra, S. R. (2021). A comprehensive review on epigenetic mechanisms and application of epigenetic modifications for crop improvement. Environ. Exp. Bot. **188:** 104479.

Sato, K., Abe, F., Mascher, M., Haberer, G., Gundlach, H., Spannagl, M., Shirasawa, K., and Isobe, S. (2021). Chromosome-scale genome assembly of the transformation-amenable common wheat cultivar 'Fielder'. DNA Res. **28:** dsab008.

Scossa, F., Alseekh, S., and Fernie, A.R. (2021). Integrating multi-omics data for crop improvement. J. Plant Physiol. **257:** 153352.

Scott, M.F., Fradgley, N., Bentley, A.R., Brabbs, T., Corke, F., Gardner, K.A., Horsnell, R., Howell, P., Ladejobi, O., Mackay, I.J., et al. (2021). Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. Genome Biol. **22:** 137.

Shewry, P.R., and Hey, S.J. (2015). The contribution of wheat to human diet and health. Food Energy Secur. **4:** 178-202.

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

Shi, X., Cui, F., Han, X., He, Y., Zhao, L., Zhang, N., Zhang, H., Zhu, H., Liu, Z., Ma, B., et al. (2022). Comparative genomic and transcriptomic analyses uncover the molecular basis of high nitrogen-use efficiency in the wheat cultivar Kenong 9204. Mol. Plant **15:** 1440-1456.

Shimizu, K.K., Copetti, D., Okada, M., Wicker, T., Tameshige, T., Hatakeyama, M., Shimizu-Inatsugi, R., Aquino, C., Nishimura, K., Kobayashi, F., et al. (2021). *De novo* genome assembly of the Japanese wheat cultivar Norin 61 highlights functional variation in flowering time and *Fusarium*-resistant genes in East Asian genotypes. Plant Cell Physiol. **62:** 8-27.

Song, Q., Zhang, T., Stelly, D.M., and Chen, Z.J. (2017). Epigenomic and functional

analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol. **18:** 99.

Soyk, S., Benoit, M., and Lippman, Z.B. (2020). New horizons for dissecting epistasis in crop quantitative trait variation. In Annu. Rev. Genet., Vol 54, 2020, N.M. Bonini, ed. pp. 287-307.

Strejčkova, B., Milec, Z., Holusova, K., Capal, P., Vojtkova, T., Cegan, R., and Safar, J. (2021). In-depth sequence analysis of bread wheat *VRN1* genes. Int. J. Mol. Sci. **22:** 12284.

Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., and Chen, F. (2020). The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnol J. **18:** 1354-1360.

Sun, Y., Shang, L., Zhu, Q.H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. Trends Plant Sci. **27:** 391-401.

Tibbs Cortes, L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. Plant Genome **14:** e20077.

Tu, Y., Liu, H., Liu, J., Tang, H., Mu, Y., Deng, M., Jiang, Q., Liu, Y., Chen, G., Wang, J., et al. (2021). QTL mapping and validation of bread wheat flag leaf morphology across multiple environments in different genetic backgrounds. Theor. Appl. Genet. **134:** 261-278.

Varshney, R.K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M.E. (2021). Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci. **26:** 631-649.

Venske, E., dos Santos, R.S., Busanello, C., Gustafson, P., and de Oliveira, A.C. (2019). Bread wheat: a role model for plant domestication and breeding. Hereditas **156:** 16.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H., Kolodziej, M.C., Delorean, E., Thambugala, D., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature **588:** 277-283.

Wang, D., Li, F., Cao, S., and Zhang, K. (2020). Genomic and functional genomics analyses of gluten proteins and prospect for simultaneous improvement of end-use and health-related traits in wheat. Theor. Appl. Genet. **133:** 1521-1539.

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S., et al. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. Nat. Genet. **44:**

1098-1103.

Wang, M., Li, Z., Zhang, Y., Zhang, Y., Xie, Y., Ye, L., Zhuang, Y., Lin, K., Zhao, F., Guo, J., et al. (2021). An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. Plant Cell **33:** 865-881.

Wang, Y., Shi, C., Yang, T., Zhao, L., Chen, J., Zhang, N., Ren, Y., Tang, G., Cui, D., and Chen, F. (2018). High-throughput sequencing revealed that microRNAs were involved in the development of superior and inferior grains in bread wheat. Sci. Rep. **8:** 13854.

Wang, Z., Wang, W., Xie, X., Wang, Y., Yang, Z., Peng, H., Xin, M., Yao, Y., Hu, Z., Liu, J., et al. (2022a). Dispersed emergence and protracted domestication of polyploid wheat uncovered by mosaic ancestral haploblock inference. Nat. Commun. **13:** 3891.

Wang, Z., Zhao, G., Yang, Q., Gao, L., Liu, C., Ru, Z., Wang, D., Jia, J., and Cui, D. (2022b). Helitron and CACTA DNA transposons actively reshape the common wheat-AK58 genome. Genomics **114:** 110288.

Weckwerth, W., Ghatak, A., Bellaire, A., Chaturvedi, P., and Varshney, R.K. (2020). PANOMICS meets germplasm. Plant Biotechnol. J. **18:** 1507-1525.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., and Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA **106:** 13875-13879.

Wu, K., Wang, S., Song, W., Zhang, J., Wang, Y., Liu, Q., Yu, J., Ye, Y., Li, S., Chen, J., et al. (2020). Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. Science **367:** eaaz2046.

Xiao, J., Liu, B., Yao, Y., Guo, Z., Jia, H., Kong, L., Zhang, A., Ma, W., Ni, Z., Xu, S., et al. (2022). Wheat genomic study for genetic improvement of traits in China. Sci. China Life Sci. **65:** 1718-1775.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S., and Dubcovsky, J. (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of FT. Proc. Natl. Acad. Sci. USA. **103:** 19581-6.

Yang, Y., Zhang, X., Wu, L., Zhang, L., Liu, G., Xia, C., Liu, X., and Kong, X. (2021). Transcriptome profiling of developing leaf and shoot apices to reveal the molecular mechanism and co-expression genes responsible for the wheat heading date. BMC Genomics

22: 468.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203-208.

Yuan, J., Jiao, W., Liu, Y., Ye, W., Wang, X., Liu, B., Song, Q., and Chen, Z.J. (2020). Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. BMC Biol. **18:** 171.

Yuan, J., Sun, H., Wang, Y., Li, L., Chen, S., Jiao, W., Jia, G., Wang, L., Mao, J., Ni, Z., et al. (2022). Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. Genome Biol. **23:** 34.

Yuan, Y., Bayer, P.E., Batley, J., and Edwards, D. (2021). Current status of structural variation studies in plants. Plant Biotechnol. J. **19:** 2153-2163.

Zhai, H., Jiang, C., Zhao, Y., Yang, S., Li, Y., Yan, K., Wu, S., Luo, B., Du, Y., Jin, H., et al. (2021). Wheat heat tolerance is impaired by heightened deletions in the distal end of 4AL chromosomal arm. Plant Biotechnol. J. **19:** 1038-1051.

Zhang, K., Wang, J., Zhang, L., Rong, C., Zhao, F., Peng, T., Li, H., Cheng, D., Liu, X., Qin, H., et al. (2013). Association analysis of genomic loci important for grain weight control in elite common wheat varieties cultivated with variable water and fertiliser supply. PLoS One **8:** e57853.

Zhang, X., Wang, H., Sun, H., Li, Y., Feng, Y., Jiao, C., Li, M., Song, X., Wang, T., Wang, Z., et al. (2023). A chromosome-scale genome assembly of *Dasypyrum villosum* provides insights into its application as a broad-spectrum disease resistance resource for wheat improvement. Mol. Plant **16:** 432-451.

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., Liu, J., Chen, K., Zhang, H., Gao C., et al. (2018). Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. Plant J. **94:** 857-866.

Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J., Huneau, C., Magdelenat, G., Couloux, A., Samain, S., Gill, B., et al. (2011). Duplication and partitioning in evolution and function of homoeologous *Q* loci governing domestication characters in polyploid wheat.

Proc. Natl. Acad. Sci. USA **108:** 18737-18742.

Zhao, C., Bao, Y., Wang, X., Yu, H., Ding, A., Guan, C., Cui, J., Wu, Y., Sun, H., Li, X., et al. (2018). QTL for flag leaf size and their influence on yield-related traits in wheat. Euphytica **214:** 209.

Zhao, J., Xie, Y., Kong, C., Lu, Z., Jia, H., Ma, Z., Zhang, Y., Cui, D., Ru, Z., Wang, Y., et al. (2023a). Centromere repositioning and shifts in wheat evolution. Plant Commun. 100556.

Zhao, L., Xie, L., Zhang, Q., Ouyang, W., Deng, L., Guan, P., Ma, M., Li, Y., Zhang, Y., Xiao, Q., et al. (2020). Integrative analysis of reference epigenomes in 20 rice varieties. Nat. Commun. **11(1):** 2658.

Zhao, X., Guo, Y., Kang, L., Yin, C., Bi, A., Xu, D., Zhang, Z., Zhang, J., Yang, X., Xu, J., et al. (2023b). Population genomics unravels the Holocene history of bread wheat and its relatives. Nat. Plants **9:** 403-419.

Zheng, Y., Jiao, C., Sun, H., Rosli, Hernan G., Pombo, Marina A., Zhang, P., Banf, M., Dai, X., Martin, Gregory B., Giovannoni, James J., et al. (2016). iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol. Plant **9:** 1667-1670.

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y.G., et al. (2020). *Triticum* population sequencing provides insights into wheat adaptation. Nat. Genet. **52:** 1412-1422.

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M.K., Zhang, C., Chang, W.C., Zhang, L., Zhang, X., Tang, R., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat. Genet. **51:** 865-876.

Figure legends

Figure 1. **Main features of AK58 genome assembly.**

An outline of AK58 genomic features. **Track a,** the 21 chromosomes. One scale label indicates 10 Mb. The black histogram indicates the distribution of two types of LTR-RTs (*Quinta* and *Cereba*), with the peaks indicating candidate centromeric

1228 regions. **Track b,** Distribution of the 105 known miRNAs (represented by yellow dots)
1229 on different chromosomes. **Track c,** lncRNA density, presented by lncRNA length/5
1230 Mb. **Track d,** gene density, measured by genes/5 Mb. **Track e,** gene expression,
1231 calculated as the average RPKM value per 5 Mb. **Track f,** SNP density of AK58 (as
1232 compared to CS). **Tracks g-j**, density of total TE (**g**), *Gypsy* (**h**), *Copia* (**i**), and DNA
1233 (**j**) TEs, all calculated as total length of TEs per 5 Mb.

1234

Figure 2. Analysis of the co-expression networks and modules of AK58.

(A) The 84 co-expression modules constructed based on the transcriptomic data obtained in this study for root, young leaf, flag leaf, stem, young spike or grain tissues. Heatmap of the expression pattern of a representative gene (eigengene) in the given module was defined by WGCNA. An eigengene summarizes the expression profiles of a group of co-expressed genes. Rows and columns indicate samples and modules, respectively. White boxes on the left indicate tissue types.

(B) The dynamics of TF gene expression patterns in four co-expression networks constructed using the transcriptomic data of leaf, root, grain, or young spike tissues. Node coloring is according to the clustering of co-expression modules. Putative functions for some of the genes in the network are annotated based on their orthologs characterized in rice. Gene expression networks and their modules were mainly tissue dependent instead of subgenome dependent.

(C) Comparisons of chromatin states between AK58 subgenomes and the D genome of *Ae. tauschii*. Chromatin states were obtained for AK58 subgenomes and *Ae. tauschii* genome, respectively, using a 15-state ChromHMM model based on 18 histone marks. Darker blue color in the heatmaps indicates a higher probability or enrichment of epi-marks. Rows of the heatmap correspond to the determined states, and columns correspond to different histone marks with two replicates. The states are reordered by their similarity among the four genomes.

Figure 3. Comparison of main agronmic and genomic charateristics between AK58 and CS.

(A) Plant architectures of AK58 and CS. The two cultivars differ clearly in plant height, spike, grain, leaf, and tiller angles. For each trait, AK58 is shown on the left and CS on the right.

(B) Synteny between the B subgenomes of AK58 and CS, with colinear regions connected by vertical lines.

(C) The distribution of indels along chromosome 2B (Chr2B) of AK58. Indel density is calculated in 5 Mb windows along the chromosome. The light orange bar indicates the centromeric region.

(D) Volcano plots of differentially expressed genes in AK58 compared with CS in four samples. Grain-DAF4, developing grains collected at 4 days after anthesis; FM, floret meristems at about 1cm inflorescence stage; leaf and root at seedling stage.

Figure 4. **Detection of agronomically important homoeologous loci by HGWAS.**

(A) A diagram illustrating the difference between GWAS and HGWAS approaches in common wheat. During genotyping, GWAS considers homoeologs independently for detecting the homoeologs associated with different agronomic traits, whereas HGWAS treats the three homoeologs as one genetic unit for detecting the homoeologous loci linked to specific traits. For example, the red circles in the left panel indicates trait-associating homoeologs revealed by GWAS, while in the right panel the trait-associating homoeologous loci uncovered by HGWAS are boxed in red.

(B) The distribution of 139 major HGWAS loci along the seven groups of homoeologous chromosomes (G1 - G7) detected in this study using AK58/CS F2 population. The approximate physical position (Mb) is privided on the left. Twenty traits were examined, including AL (awn length), Chl (chlorophyll content), FLL (flag leaf length), FLN (floret number per spike), FLT (flag leaf thickness), FLW (flag leaf width), GL (grain length), GN (grain number per spike), GW (grain width), HT (heading time), LA (leaf angle), MT (mature time), PH (plant height), Pm (powdery mildew resistance), SD (spikelet density), SL (spike length), SLN (spikelet number per spike), SN (spike number per plant), SS (seed setting), TGW (1000-grain weight), as listed in Supplemental Table 22.

(C) An example showing the superior efficiency of HGWAS over GWAS in detecting trait-associating chromosomal loci. The loci significantly associated with heading time were both identified by GWAS (up) and HGWAS (bottom) on group 2 and 5 chromosomes, with higher R^2 values by HGWAS. Additionally, one locus on group 1 chromosomes was detected by HGWAS but not GWAS, which explained 16% of the heading time variation.

1295 **Table 1. Summary of AK58 genome assembly**

Parameter	Length (bp)		Number	
	Contig	**Scaffold**	**Contig**	**Scaffold**
Total	14,659,748,929	14,752,721,585	279,861	159,139
Maximum	2,084,420	115,914,924	-	-
N50	237,187	28,282,379	18,423	153
N60	189,861	21,419,151	25,335	213
N70	147,027	16,464,209	34,106	290
N80	105,449	10,989,250	45,816	400
N90	59,822	5,727,441	63,842	584

1296

1297 **Table 2. Comparison of genetic effects on plot yield for the haplotypes of *Vrn3* revealed by conventional GWAS or HGWAS**

GWAS				**HGWAS**			
Associated homoeolog	**Homoeolog haplotype**	**Plot yield (kg, mean ± SD)**	**Effect of elite haplotype**	**Associated locus**	**Homoeologous haplotype**	**Plot yield (kg, mean ± SD)**	**Effect of elite haplotype**
Vrn3-7D	**Vrn3-7D-hap1**	**2.6 ± 0.7 (153) ***	**13.0%**	*Vrn3*	**HH1, Vrn3-7A-hap1_7B-hap1_7D-hap1**	**2.7 ± 0.6 (70)** ab	**17.4%**
	Vrn3-7D-hap2	1.8 ± 0.8 (114)			HH2, Vrn3-7A-hap2_7B-hap1_7D-hap1	2.5 ± 0.7 (55) abc	
					HH3, Vrn3-7A-hap1_7B-hap1_7D-hap2	2.1 ± 0.7 (38) bc	
					HH4, Vrn3-7A-hap2_7B-hap1_7D-hap2	2.0 ± 0.7 (38) bc	
					HH5, Vrn3-7A-hap3_7B-hap1_7D-hap2	1.2 ± 0.4 (34) d	
					HH6, Vrn3-7A-hap2_7B-hap2_7D-hap1	**3.0 ± 0.7 (13)** a	**30.4%**
					HH7, Vrn3-7A-hap1_7B-hap2_7D-hap1	**3.1 ± 0.6 (10)** a	**34.8%**
					HH8, Vrn3-7A-hap1_7B-hap2_7D-hap2	1.8 ± 0.6 (4) c	
					HH9, Vrn3-7A-hap3_7B-hap1_7D-hap1	1.9 ± 0.5 (5) c	

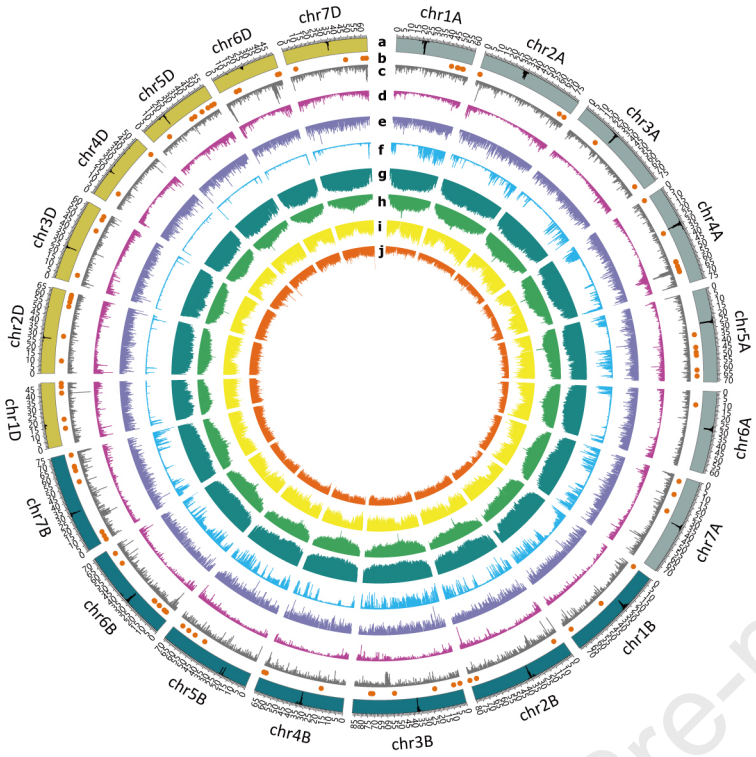
1298 The varietal population was phenotypically assessed by Gao et al. (2017). The mean plot (4.5 m^2 each) yield for the varietal population was 2.3 ± 0.8 kg. The number in the brackets indicates
1299 the lines having the given genotype. Statistical analysis was conducted using Student *t* test (for GWAS) or the LSD method with significant differences marked by different small letters after the
1300 means. In both cases, significant differences were based on $P < 0.05$.

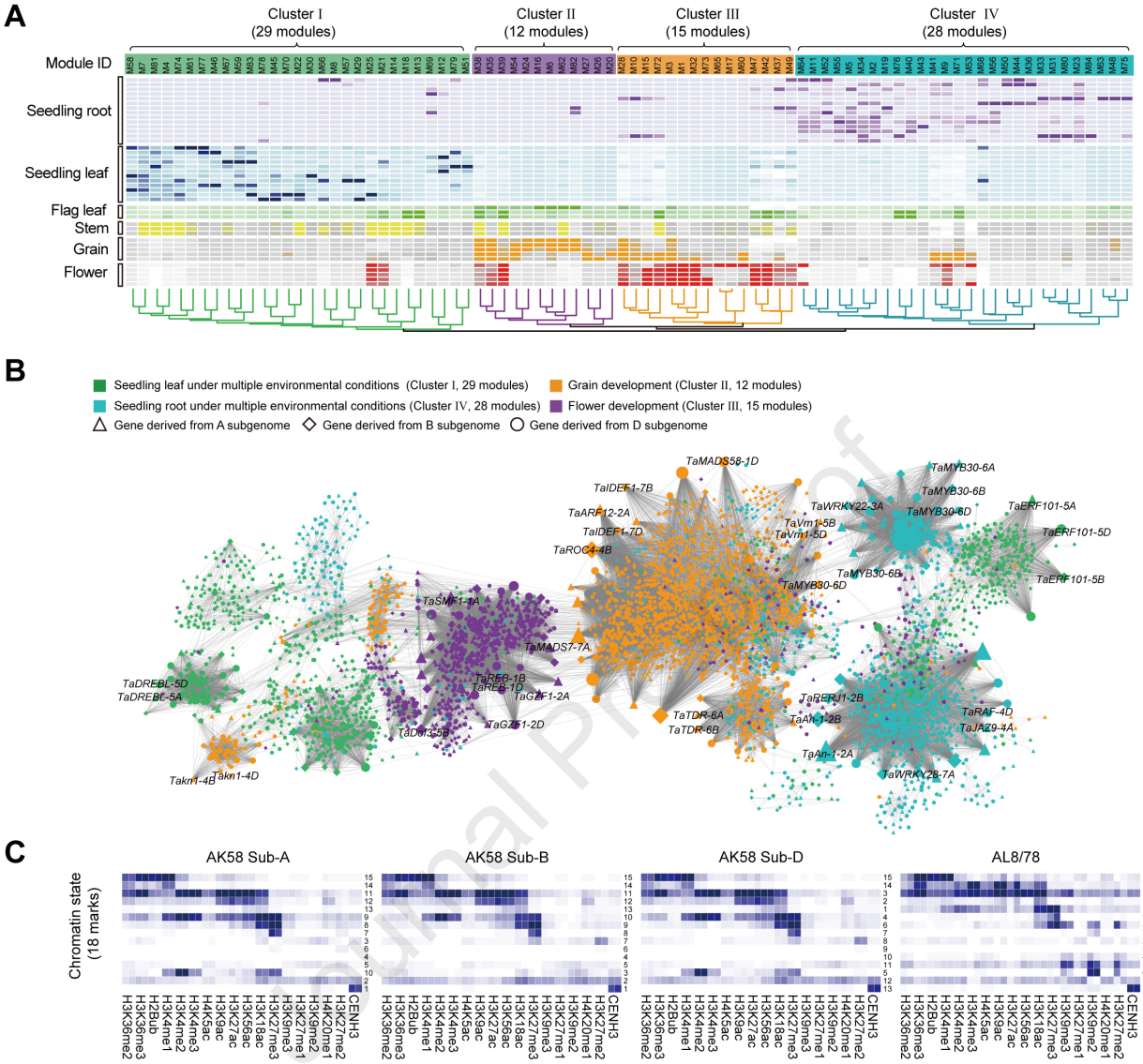
Table 3. The HGWAS loci detected in the *Vrn1* region and their elite HH haplotypes

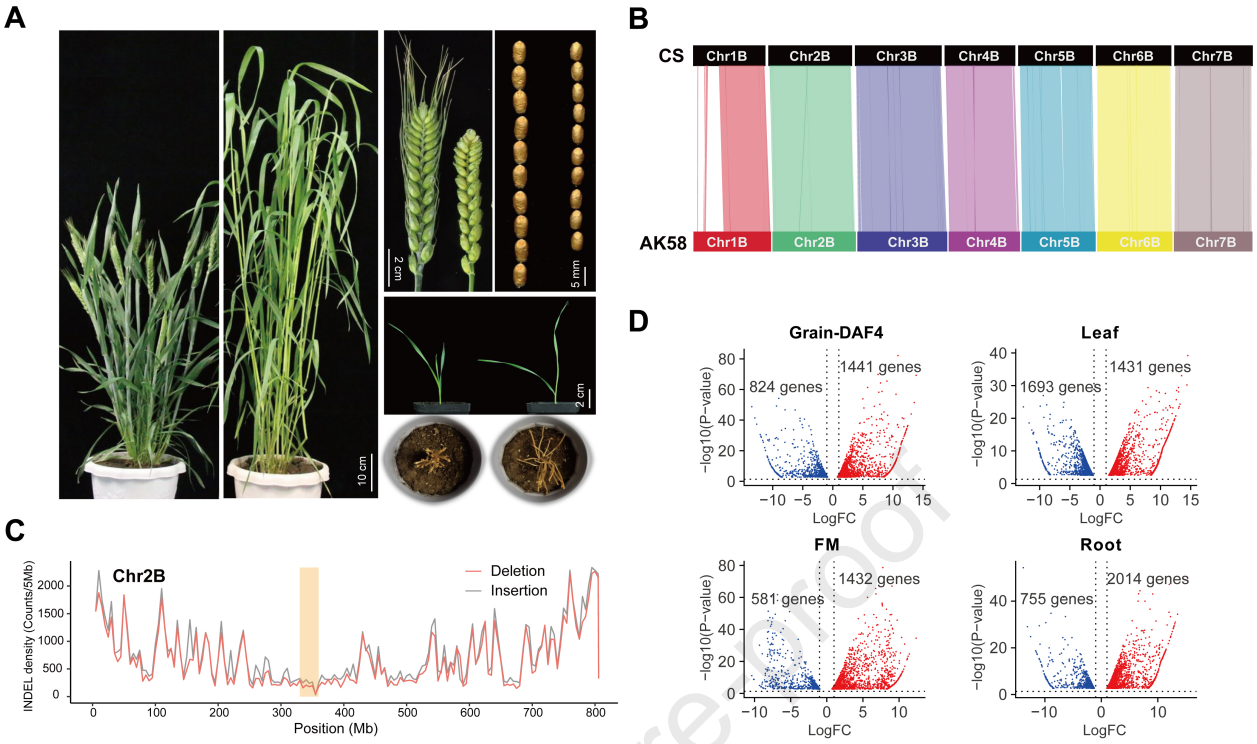
HGWAS locus	Trait concerned	AK58 trait value (HH: AAA)	CS trait value (HH: CCC)	Middle parent value (MPV)	Elite HH	Trait value of elite HH
Chl_G5_465.2_485.8	Chlorophyll content (Chl, SPAD)	50.6 ± 2.8a	51.5 ± 2.9b	51.5 ± 3.4b	CAC	53.1 ± 3.4c
FLL_G5_474.7_486.2	Flag leaf length (FLL, cm)	19.3 ± 4.7b	19.3 ± 3.1b	18. 6± 4a	AAC	17.3 ± 3.6a
FLT_G5_478.7_486.2	Flag leaf thickness (FLT, mm)	0.217 ± 0.026a	0.214 ± 0.017a	0.221 ± 0.023a	AHC	0.234 ± 0.023b
FLN_G5_465.5_492.6	Floret number per spike (FLN)	86.7 ± 15.7a	98.3 ± 14.1c	93.7 ± 17.3b	CAA	102 ± 17.7c
GN_G5_476.7_492.6	Grain number per spike (GN)	52.0 ± 17.7a	61.2 ± 14.5b	59.1 ± 18.6a	CAA	65.9 ± 19.9b
HT_G5_473.4_490.5	Heading time (HT, day)	191.9 ± 7.6c	184.1 ± 3.9a	188.4 ± 9.1b	AAC	182.9 ± 10.3a
PH_G5_480.1_489.9	Plant height (PH, cm)	100.8 ± 18.5c	98.6 ± 12.4b	97.4 ± 15.1b	AAC	88.7 ± 11.4a
SL_G5_473.7_489.9	Spike length (SL, cm)	8.4 ± 1.5a	8.0 ± 1.0a	8.7 ± 1.6b	CHA	9.4 ± 1.6c
SLN_G5_465.8_491.1	Spikelet number per spike (SLN)	22.4 ± 2.1b	22.1 ± 2.2a	22.4 ± 2.6b	HCA	24.2 ± 2.6c

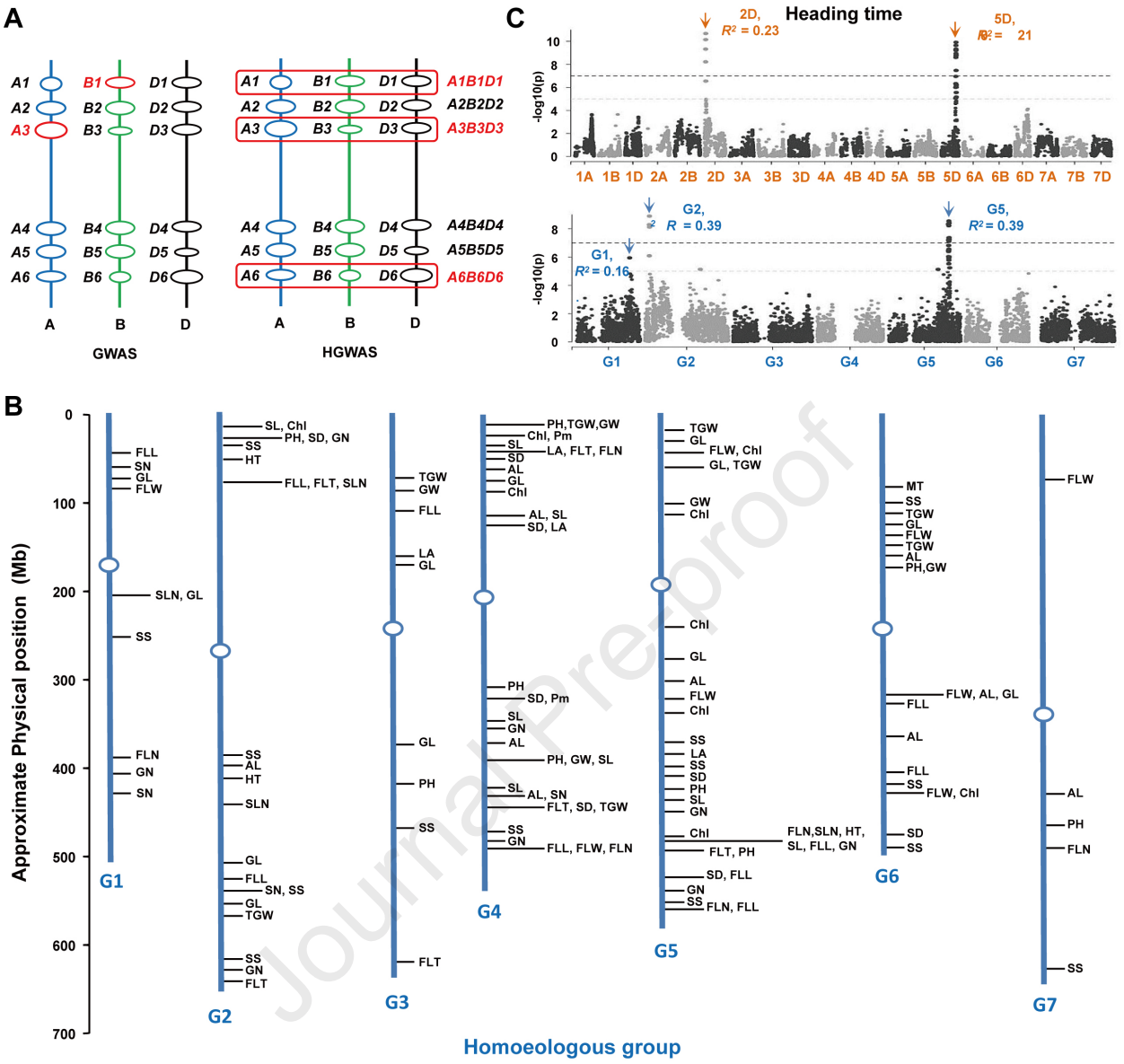
Trait values are Means ± SD. Multiple statistical comparisons were conducted using the LSD method, with different small letters indicating significant differences ($P \leq 0.05$).

1301









The declaration of interests for

Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement

Jizeng Jia[1,2,12], Guangyao Zhao[2,12], Danping Li[2,12], Kai Wang[4,12], Chuizheng Kong[2,12], Pingchuan Deng[5,11], Xueqing Yan[6,7], Xueyong Zhang[2], Zefu Lu[2], Shujuan Xu[7,8], Yuannian Jiao[6,7], Kang Chong[7,8], Xu Liu[2], Dangqun Cui[1], Guangwei Li[1], Yijing Zhang[9], Chunguang Du[1], Liang Wu[5,10], Tianbao Li[1,2], Dong Yan[2], Kehui Zhan[1], Feng Chen[1], Zhiyong Wang[1], Lichao Zhang[2], Xiuying Kong[2,*], Zhengang Ru[3,*], Daowen Wang[1, *], Lifeng Gao[2,*]

[1]College of Agronomy, Collaborative Innovation Center of Henan Grain Crops, State Key Laboratory of Wheat and Maize Crop Science, and Center for Crop Genome Engineering, Henan Agricultural University, Zhengzhou 450046, Henan, China

[2]State Key Laboratory of Crop Gene Resources and Breeding, the National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[3]School of Life Science and Technology, Henan Institute of Science and Technology, Xinxiang 453003, Henan, China

[4]Xi'An Shansheng Biosciences Co., Ltd., Xi'an 710000, China

[5]Zhejiang Provincial Key Laboratory of Crop Genetic Resources, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, Zhejiang, China

[6]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[7]University of Chinese Academy of Sciences, Beijing 100049, China

[8]Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[9]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, China

[10]Hainan Yazhou Bay Seed Laboratory, Hainan Institute of Zhejiang University, Sanya 562000, Hainan, China

[11]State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling 612100, Shaanxi, China

[12]These authors contributed equally.

[*]**Correspondence:** Xiuying Kong (kongxiuying@caas.cn), Zhengang Ru (rzgh58@163.com), Daowen Wang (dwwang@henau.edu.cn), Lifeng Gao (gaolifeng@caas.cn)

No conflict of interest is declared.