THE PLANT CELL

Large-Scale Biology

Whole-genome sequencing of diverse wheat accessions uncovers genetic changes during modern breeding in China and the United States

Jianqing Niu [ID],[1,2,†] Shengwei Ma [ID],[1,2,†] Shusong Zheng [ID],[2,†] Chi Zhang [ID],[3,†] Yaru Lu [ID],[2] Yaoqi Si [ID],[2] Shuiquan Tian [ID],[2,4] Xiaoli Shi [ID],[2] Xiaolin Liu [ID],[2,4] Muhammad Kashif Naeem [ID],[2] Hua Sun [ID],[2] Yafei Hu [ID],[3] Huilan Wu [ID],[2] Yan Cui [ID],[2] Chunlin Chen [ID],[2] Wenbo Long [ID],[2] Yue Zhang [ID],[2] Mengjun Gu [ID],[2] Man Cui [ID],[2] Qiao Lu [ID],[4] Wenjuan Zhou [ID],[2] Junhua Peng [ID],[5] Eduard Akhunov [ID],[6] Fei He [ID],[2,*] Shancen Zhao [ID][3,*] and Hong-Qing Ling [ID][1,2,4,*]

1 Hainan Yazhou Bay Seed Laboratory, Hainan, Sanya 572024, China
2 State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
3 BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China
4 College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
5 Huazhi Bio-tech Company Ltd., Changsha, Hunan 410125, China
6 Wheat Genetic Resources Center, Kansas State University, Manhattan, KS 66506, USA

*Author for correspondence: hqling@genetics.ac.cn (H.-Q.L.), fhe@genetics.ac.cn (F.H.), zhaoshancen@bgi.com (S.Z.)
†These authors contributed equally.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) are: Hong-Qing Ling (hqling@genetics.ac.cn) and Shusong Zheng (shusong@genetics.ac.cn).

Abstract

Breeding has dramatically changed the plant architecture of wheat (*Triticum aestivum*), resulting in the development of high-yielding varieties adapted to modern farming systems. However, how wheat breeding shaped the genomic architecture of this crop remains poorly understood. Here, we performed a comprehensive comparative analysis of a whole-genome resequencing panel of 355 common wheat accessions (representing diverse landraces and modern cultivars from China and the United States) at the phenotypic and genomic levels. The genetic diversity of modern wheat cultivars was clearly reduced compared to landraces. Consistent with these genetic changes, most phenotypes of cultivars from China and the United States were significantly altered. Of the 21 agronomic traits investigated, 8 showed convergent changes between the 2 countries. Moreover, of the 207 loci associated with these 21 traits, more than half overlapped with genomic regions that showed evidence of selection. The distribution of selected loci between the Chinese and American cultivars suggests that breeding for increased productivity in these 2 regions was accomplished by pyramiding both shared and region-specific variants. This work provides a framework to understand the genetic architecture of the adaptation of wheat to diverse agricultural production environments, as well as guidelines for optimizing breeding strategies to design better wheat varieties.

Introduction

Common wheat (*Triticum aestivum*) spread to a wide range of diverse environments since it originated in the Fertile Crescent approximately 8,000 years ago (Marcussen et al.

2014). During this process, wheat experienced intensive natural and artificial selection associated with adaptation to new environments, human needs, and local agricultural practices, resulting in the development of local landraces. A landrace is

IN A NUTSHELL

Background: A landrace is a traditional plant variety capable of tolerating local biotic and abiotic stresses and maintaining intermediate yield levels in low-input farming systems. The development of cultivars from landraces was achieved by human-mediated selection aimed at higher yield, better quality, and stronger fitness. The transition from landraces to elite cultivars is regarded as crop improvement, during which beneficial alleles might gradually accumulate in cultivars. Therefore, understanding the genetic architecture of wheat (*Triticum aestivum*) improvement during the transition from landraces to cultivars in distinct geographic regions will be crucial for developing high-performance varieties in the future.

Question: How has modern wheat breeding reshaped the phenotypic and genomic architecture of wheat in China and the United States?

Findings: We performed a comprehensive comparative analysis (at the phenotypic and genomic levels) of a whole-genome resequencing panel of 355 common wheat accessions representing diverse landraces and modern cultivars from China and the United States. Compared with landraces, the genetic diversity and phenotypes of modern wheat cultivars from China and the United States changed significantly. Furthermore, we identified breeding targets during modern wheat breeding and determined that breeding for increased productivity in these 2 geographic regions was accomplished by pyramiding shared and region-specific variants.

Next steps: The unique loci selected either in China or in the United States can be used to develop high-performance wheat varieties in the future.

a traditional crop variety capable of tolerating local biotic and abiotic stresses and maintaining intermediate yield levels in low-input farming systems (Zeven 1998; Lopes et al. 2015). The development of cultivars from landraces was achieved by human-mediated selection aimed at higher yield, better quality, and stronger fitness (Liu et al. 2019). However, most modern cultivars were developed from a limited number of founders, resulting in a bottleneck effect and reduced genetic diversity in most breeding programs (Lopes et al. 2015). This hinders the progress toward the development of new wheat varieties adapted to more extreme environmental conditions to provide food security for the growing human population (McCouch et al. 2013; Bhatta et al. 2018). Wheat landraces, a rich reservoir of regional adaptive diversity, represent valuable genetic resources for overcoming these challenges (Milner et al. 2019). Therefore, understanding the genetic architecture of wheat improvement during the transition from landraces to cultivars (Cavanagh et al. 2013; Zhou et al. 2018) in distinct geographic regions will be crucial for developing high-performance varieties in the future.

Comprehensive characterization of genetic variations across the genome, identification of loci subjected to selection during breeding, and mapping markers associated with variation in agronomic traits can provide guidelines for further wheat improvement and can help improve genomic selection models (Morrell et al. 2011; Cavanagh et al. 2013). In the past few years, studies of the history, genomic composition, gene flow, selective sweeps, and genetic basis of wheat have been greatly promoted by population genomics. Common wheat originated from the southwest coast of the Caspian Sea and spread across Eurasia and reached

Europe, South Asia, and East Asia. This was accompanied by frequent intra- and interspecies introgression from wheat relatives to facilitate its proliferation in novel environments (Cheng et al. 2019; He et al. 2019; Pont et al. 2019; Guo et al. 2020; Zhou et al. 2020; Wang et al. 2022; Zhao et al. 2023).

Following adaptation to local environments, landraces emerged and underwent a small population bottleneck and successive selection during modern wheat breeding to develop cultivars (Cavanagh et al. 2013). During this process, some genetic footprints related to grain yield, growth periods, disease resistance, vernalization, and flour quality were selected (Gaire et al. 2020; Hao et al. 2020; Sansaloni et al. 2020; Li et al. 2022). The selective genetic footprints of modern wheat varied through time and among different breeding programs. Most studies investigating this process have focused on a single breeding program using either populations of relatively small size or with low marker density. The relatively small sizes of the studied populations, modest marker density, or the low number of analyzed traits only provided limited insights into the genetic impact of trait improvement, especially in multiple regional breeding programs with different management practices or growing conditions.

China and the United States are the world's major wheat producers and consumers. Although the breeding goals are the same in both counties: improving grain yield, resistance, and quality, the environmental factors and human preferences may be different. The major difference between the 2 countries is that wheat is fed with sufficient water and nutrition in China, while it is rain fed in the United States. To investigate how modern wheat breeding reshapes the plant and genome architecture in these 2 wheat-growing regions,

we sequenced 355 common wheat accessions, including cultivars from China and the United States and landraces from 13 countries, and constructed a comprehensive genomic variation map. By comparing the phenotypic changes and genomic compositions of cultivars developed in China and the United States, we uncovered similarities and differences in the direction of breeding selection between these 2 countries. Furthermore, we constructed genome-wide maps of selective sweeps and phenome-to-genome associations, which incorporate 207 loci linked with 21 key agronomic traits. Our study provides insights into the impacts of improvement selection in both China and the United States on the genetic architecture of major agronomic traits in common wheat. We also generated resources for further wheat improvement.

Results

Whole-genome resequencing of 355 common wheat germplasms reveals abundant genetic diversity

To identify genomic regions targeted by modern wheat breeding in China and the United States, we analyzed 355 common wheat accessions, including 175 improved cultivars (103 from China and 72 from the United States) and 180 representative diverse landraces from 13 countries mainly around the Fertile Crescent (Supplemental Data Set 1). Most of the cultivars were collected from the major wheat production areas in China and the United States. The landraces harbor more than 94% of the diversity ($\pi_{180\ Landraces}/\pi_{632\ Landraces}$) of a previously described population of 632 worldwide wheat landraces (Balfourier et al. 2019), and all 632 landraces match to at least 1 counterpart in the 180 landraces examined in the current study (Supplemental Fig. S1). The 355 common wheat accessions were collected from 14 countries (Supplemental Data Set 2), representing geographically and genetically diverse landraces and cultivars from distinct climatic regions.

Whole-genome sequencing of the 355 wheat accessions yielded 799.58 billion 100-bp paired-end reads with an average genome coverage depth of 14.52× for each accession (Supplemental Data Set 3). By mapping the quality-trimmed reads to the wheat reference genome (IWGSC RefSeq v1.0) and implementing a strict filtering pipeline, we ultimately identified 76,874,471 high-quality single nucleotide polymorphisms (SNPs) and 5,208,800 insertions and deletions (InDels ≤ 8 bp) (Supplemental Data Set 4). To estimate the accuracy of the SNPs identified in this study, we also genotyped 343 of the 355 accessions using the wheat 660K SNP arrays (Sun et al. 2020) and discovered 367,846 SNPs (minor allele frequency [MAF] > 0.01, heterozygosity < 0.3, and missing rate < 0.25) in the panel, which is only 0.48% of the SNPs identified by genome resequencing. We used the SNPs shared between the genome resequencing data and the wheat 660K SNP array in each accession to evaluate the accuracy of genotyping. The SNPs identified by whole-

genome resequencing and the wheat 660K array showed high concordance, with average of 99.74% and median of 99.91% (Supplemental Data Set 5). These results indicate that the SNPs identified by whole-genome resequencing are much more abundant (more than 220 times) than those from the wheat 660K SNP array and that the quality is high.

We analyzed the distribution of variants in common wheat. More variants occur at the ends of chromosomes than in the centromere. Chromosome 3B (chr3B) contains the most variants, whereas chr4D contains the fewest (Supplemental Data Set 4 and Fig. S2, A to D). Among the 3 subgenomes of common wheat, the B subgenome harbors the most variants, whereas the D subgenome harbors the fewest (Supplemental Fig. S2, E and F). The trends in variant distribution observed in this study are consistent with previous reports (Jordan et al. 2015; Pont et al. 2019; Hao et al. 2020). The average variant density in the whole genome was 5.64 per kb (6.28 per kb for the A subgenome, 7.72 per kb for the B subgenome, and 2.65 per kb for the D subgenome). This is by far the highest density variant map constructed in common wheat. Among these variants, 1,659,126 (2.05%) SNPs and 253,492 (4.31%) InDels are located in genic regions, covering 83.37% (89,946/107,891) of all high-confidence (HC) genes reported in the Chinese Spring genome (International Wheat Genome Sequencing Consortium (IWGSC) 2018). The number of nonsynonymous SNPs and tentatively deleterious variants (including SNPs and InDels), which potentially affect plant fitness, was 318,707 and 42,946, respectively (Supplemental Data Set 6). These deleterious variants potentially affect 18,488 HC genes, including 95 known genes such as *Rht1* and *Rht2* related to plant height (PH), *Vrn1* and *Vrn3* associated with vernalization, and multiple resistance genes (*Sr45*, *Lr21*, *Pm3*, *Pm5e*, and others).

Modern breeding has greatly shaped the genetic architecture of common wheat populations

To assess population structure among the 355 common wheat accessions, we performed genetic assignment analysis and principal component analysis (PCA) of 304,744 SNPs selected based on the patterns of linkage disequilibrium (LD). As shown in Fig. 1A, genetic assignment analysis showed that most of the cultivars from China and the United States were separated from the other accessions at K = 3. The landraces gradually separated into minor subpopulations with increasing K. PCA also suggested that most of the landraces (hereafter Landrace) clustered together, although a few accessions were close to cultivars. The cultivars were mainly grouped into 2 clusters (Fig. 1B; Supplemental Fig. S3): 1 cluster consisting of cultivars mainly collected from the United States (hereafter USA_CV) and the other mostly collected from China (hereafter CHN_CV). In addition, there was a subgroup (Mixed) including 23 accessions collected from different countries, suggesting exchange of genetic material between different countries. At K = 3, on
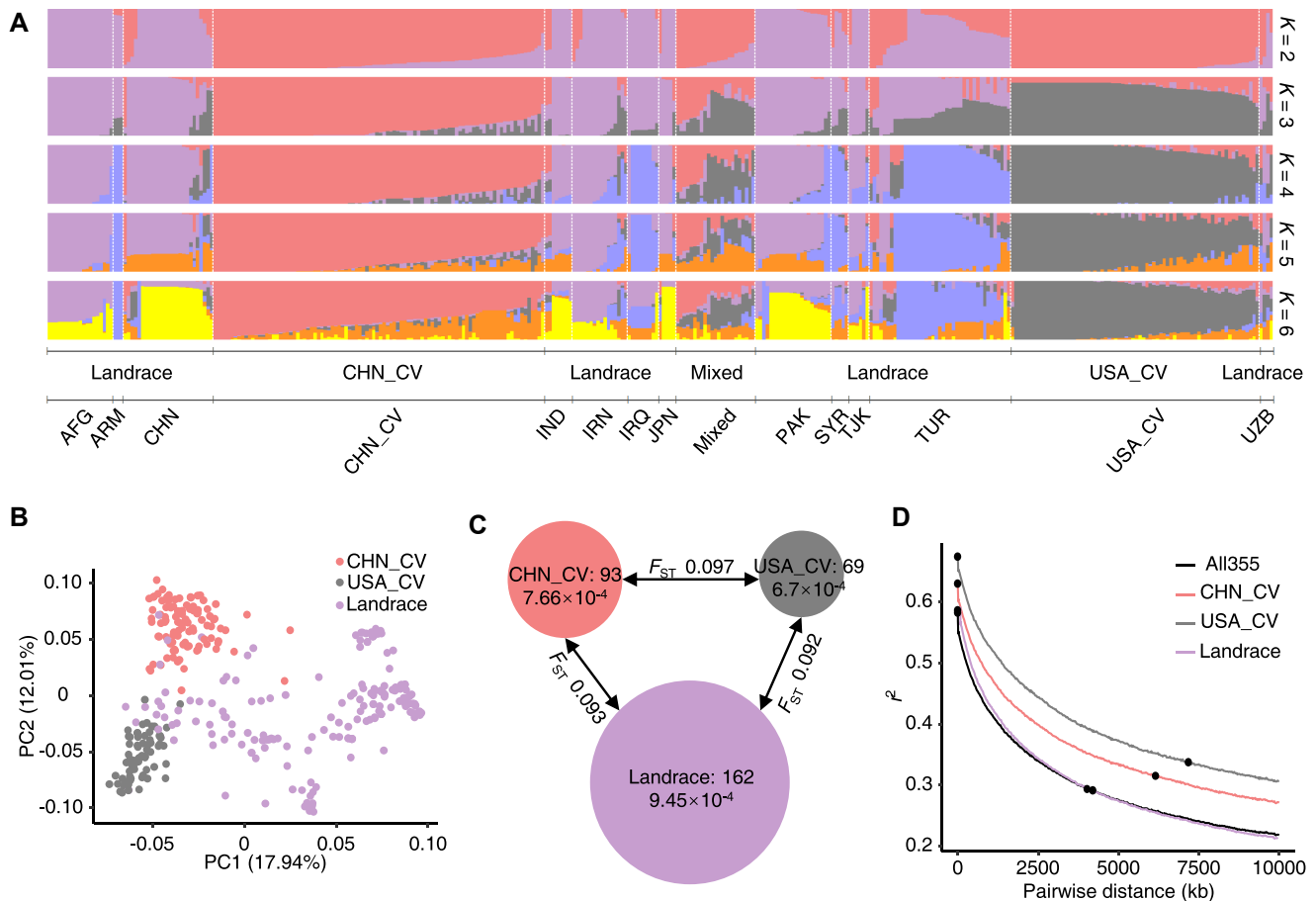
Figure 1. Population structure, genetic diversity, and LD decay of the 355 common wheat accessions investigated. **A)** Population structure of the common wheat accessions inferred using different numbers of clusters ($K = 2$ to 6). At $K = 3$, the cultivars from China and the United States were separated. Landraces were gradually separated into minor subpopulations with increasing K. AFG, Afghanistan; ARM, Armenia; CHN, China; IND, India; IRN, Iran; IRQ, Iraq; JPN, Japan; PAK, Pakistan; SYR, Syrian; TJK, Tajikistan; TUR, Turkey; UZB, Uzbekistan. **B)** PCA plot of the first 2 dimensions of the genotype data from all accessions. **C)** Statistics for genetic diversity and population differentiation. The size of the circle represents the number of accessions in each subpopulation. The numbers in the circles indicate the number of accessions and nucleotide diversity (π) in the corresponding subpopulation. The values between pairs of subpopulations show the population divergence (F_{ST}). **D)** Decay of LD in the wheat genome.

average, the Mixed group had 45.46%, 43.71%, and 10.81% of ancestry assigned to USA_CV, CHN_CV, and Landrace, respectively (Supplemental Fig. S4). Furthermore, negative f_3 statistics were obtained from f_3 (Mixed; CHN_CV and USA_CV) (Supplemental Data Set 7). These results suggest that the Mixed group was derived from at least 2 different groups related to CHN_CV and USA_CV.

It is well known that wheat originated from the Fertile Crescent and spread around the world. Therefore, landraces around this region should represent relatively raw genetic reservoirs, although new alleles likely originated during their spread worldwide. To test this hypothesis, we evaluated the number of private variants in each subpopulation (Landrace, CHN_CV, and USA_CV). The percentages of private variants were 2.76%, 3.37%, and 22.5% for CHN_CV, USA_CV, and Landrace (2,263,507/82,083,271 in CHN_CV; 2,766,702/82,083,271 in USA_CV; and 18,465,375/82,083,271 in Landrace), respectively. Furthermore, we conducted a comprehensive comparison of

genetic diversity and differentiation across populations. The genetic diversity (π) was lower in cultivars ($\pi_{CHN\_CV} = 7.66 \times 10^{-4}$ and $\pi_{USA\_CV} = 6.70 \times 10^{-4}$) than in landraces ($\pi_{Landrace} = 9.45 \times 10^{-4}$) (Fig. 1C; Supplemental Fig. S5A). Compared with Landrace, CHN_CV showed a decrease in nucleotide diversity of approximately 18.9% (2-tailed t test, $P < 1.0 \times 10^{-308}$), whereas the diversity of USA_CV showed a decrease of 29.1% (2-tailed t test, $P < 1.0 \times 10^{-308}$). These results imply that a significant bottleneck effect has arisen during modern wheat breeding.

Among the 3 subgenomes of common wheat, the nucleotide diversity of the B subgenome ($\pi_B = 1.19 \times 10^{-3}$) is the highest, that of the A subgenome ($\pi_A = 9.14 \times 10^{-4}$) is intermediate, and that of the D subgenome ($\pi_D = 2.01 \times 10^{-4}$) is the lowest (only ~10% that of the A and B subgenomes), consistent with previous studies (Cheng et al. 2019; Hao et al. 2020; Zhou et al. 2020) (Supplemental Fig. S5, B and C). The fixation index (F_{ST}) values, which describe the genetic

differentiation level between 2 populations, were 0.093 for CHN_CV vs. Landrace and 0.092 for USA_CV vs. Landrace (Fig. 1C; Supplemental Fig. S6). Both of these values are higher than that reported between improved cultivars and landraces of upland cotton (*Gossypium hirsutum*) ($F_{ST} = 0.04$) (Fang et al. 2017) and lower than that reported for soybean (*Glycine max*) ($F_{ST} > 0.106$) (Zhou et al. 2015). Notably, the genetic differentiation between CHN_CV and USA_CV ($F_{ST} = 0.097$) was higher than that between Landrace and the cultivars ($F_{ST} = 0.093$, 2-tailed t test, $P = 2.40 \times 10^{-45}$ for CHN_CV vs. Landrace and $F_{ST} = 0.092$, 2-tailed t test, $P = 1.69 \times 10^{-59}$ for USA_CV vs. Landrace), which may reflect genetic drift or different breeding strategies. The rate of LD decay (indicated by a decrease in r^2 to half of its maximum value) was 4.2 Mb for all samples, which is comparable to previous findings (Pang et al. 2020). LD decay rates varied among different populations, including 4.03 Mb for Landrace, 6.15 Mb for CHN_CV, and 7.18 Mb for USA_CV (Fig. 1D).

Modern cultivars from China and the United States have experienced different degrees of phenotypic change during wheat improvement

As mentioned above, modern cultivars have reduced genetic diversity compared to landraces (Supplemental Fig. S5A). The genetic differentiation of the cultivars from China and the United States is higher than that observed between the cultivars and landraces (Supplemental Fig. S6A). These differences could be associated with improvement selection acting on distinct genetic variants controlling traits important for agricultural productivity in different environments. One of the major targets of breeding is to increase productivity, which can be achieved by altering distinct combinations of component traits (e.g. grain size, grain number, and number of spikes) with potential to positively affect yield. To test these hypotheses, we phenotyped our panel for 21 key agronomic traits under the same culture condition in 3 independent replicates for 3 consecutive years (Supplemental Data Set 8). All 21 key agronomic traits showed significant differences (2-tailed t test, $P < 0.01$) between CHN_CV and USA_CV (Fig. 2; Supplemental Fig. S7 and Data Set 9). Furthermore, we observed convergent phenotypic changes in plant architecture and grain yield components in both CHN_CV and USA_CV compared with Landrace (Fig. 2; Supplemental Fig. S7 and Data Set 9), which was associated with higher yield per plant (YPP), biomass per plant (BPP), and harvest index (HI). However, the extent of phenotypic changes differed between China and the United States. For example, PH, peduncle length (PL), spike length (SPL), and flag leaf length (FLL) were significantly reduced in both countries compared to landraces. However, the reductions in these 4 traits were clearly greater in China (27.61% to 48.72%) than in the United States (13.38% to 22.61%) (Fig. 2; Supplemental Fig. S7 and Data Set 9). These results imply that genetic changes during modern breeding for these traits in China and the United States have been similar in direction but different in magnitude.

Identification of candidate genes and loci for multiple agronomic traits by genome-wide association study

To characterize the genetic basis underlying phenotypic changes in the American and Chinese breeding programs, we performed genome-wide association study (GWAS) of the 21 important agronomic traits and identified 5,931 marker–trait associations (MTAs; Supplemental Data Set 10) assigned to 207 loci with a suggestive threshold ($P < 1 \times 10^{-6}$; false discovery rate [FDR] < 0.05; Supplemental Data Set 11). Among these associated loci, 6 are located in genomic regions harboring known genes, such as *Reduced height-2* (*Rht2*) (Peng et al. 1999) for PH and PL, *FRIZZY PANICLE* (*WFZP-A*) (Du et al. 2021) for SN, *PHOTOPERIOD 1* (*PPD1*) (Turner et al. 2005) for anthesis days (AD), heading days (HD), PH, and PL, *Tipped 1* (*B1/ALI-1*) (DeWitt et al. 2019; Huang et al. 2019; Wang et al. 2019; Niu et al. 2020) for awn length (AL), *WHEAT ORTHOLOG OF APO 1* (*WAPO1*) (Kuzay et al. 2019; Muqaddasi et al. 2019) for spikelet number (SN), and *VERNALIZATION 3* (*VRN3*) (Yan et al. 2006) for AD and HD. These results validate the results of GWAS (Supplemental Fig. S8). We also identified 10 associations (3 for seed length [SL] and 1 each for tiller number during the mature stage [TNMS], seed width [SW], thousand kernel weight [TKW], HD, AD, sterile spikelet number [SSN], and SPL) located within the candidate genes carrying polymorphisms affecting gene function.

The rice (*Oryza sativa*) homologs of all these wheat candidate genes are responsible for similar traits. For example, 2 genes were associated with SL: *TaAKT2*, the homolog of the gene encoding a rice shaker potassium channel protein involved in regulating grain shape through the redistribution of K^+, and *TaPK4*, encoding a mitochondria-associated pyruvate kinase that regulates grain filling (Hu et al. 2020; Tian et al. 2021) (Fig. 3, A to D). In addition, *TaWTG1/TaOTUB1*, which encodes a human OTUB1-like deubiquitinase that influences grain size and shape, was simultaneously associated with SL and SPL (Huang et al. 2017; Wang et al. 2017) (Fig. 3, A and B; Supplemental Fig. S9, A and E). Interestingly, we also identified *TaOFP14* and *TaGS9*, 2 promising candidate genes for grain shape and TKW, whose corresponding homologs *OsOFP14* and *OsGS9* interact with each other to regulate grain shape and grain yield in rice (Zhao et al. 2018) (Fig. 3, A and D; Supplemental Fig. S9, D and H). Consistent with the documented roles of the corresponding rice homologs, *TaIAGLU* (homologous to rice *OsIAGLU*, which encodes an indole-3-acetic acid–conjugating enzyme), *TaPIL13* (homologous to the atypical HLH gene *OsPIL13*), and *TaRDR6* (homologous to rice *RNA-DEPENDENT POLYMERASE 6*) were associated with TNMS, HD, and SSN, respectively (Zhao et al. 2011; Song et al. 2012; Choi et al. 2013) (Fig. 3; Supplemental Fig. S9).

For each of these candidate genes, we performed haplotype-based association analysis and demonstrated that haplotypes carrying putative polymorphisms causing missense and/or frameshift mutations in the corresponding gene(s) are significantly associated with phenotypic changes.
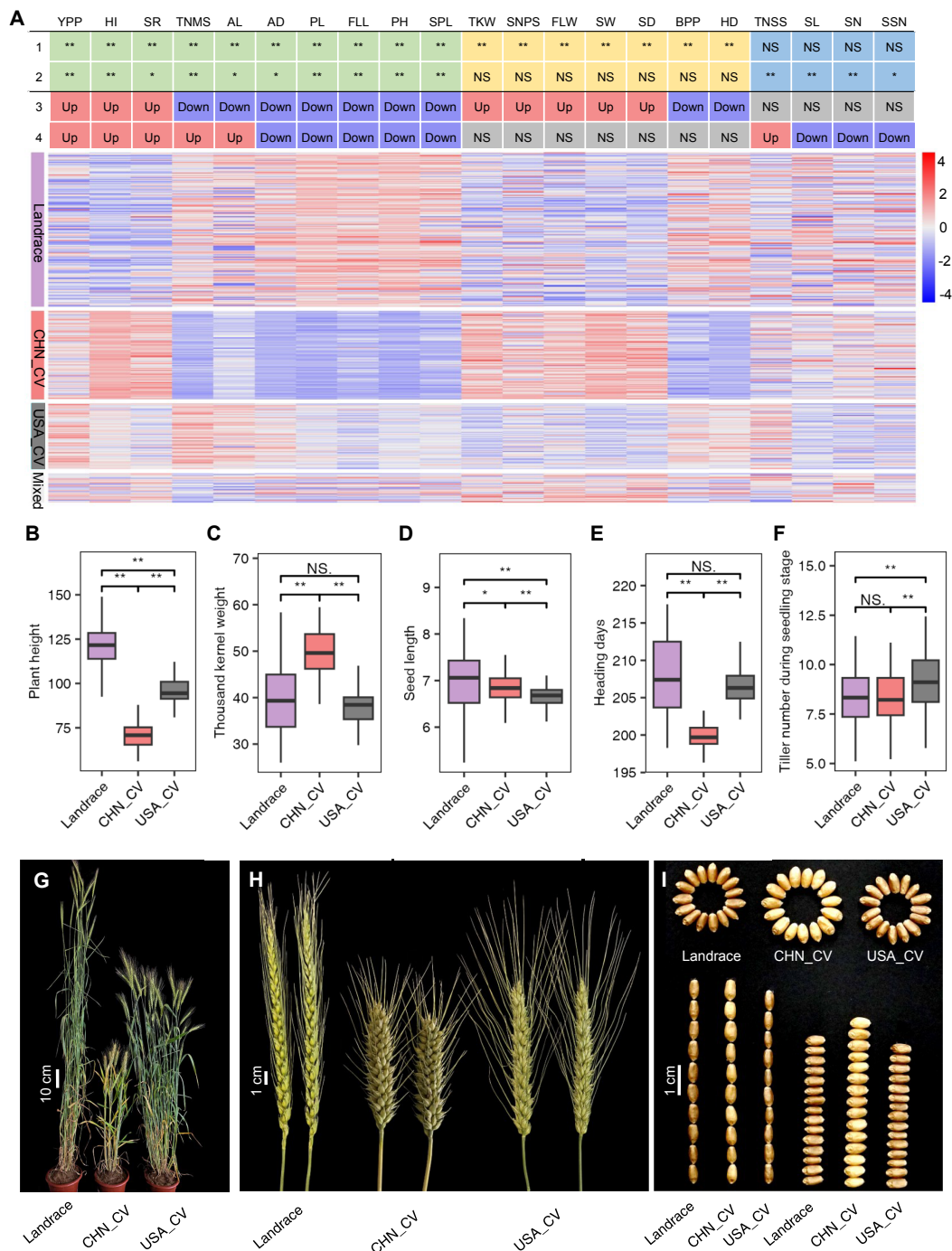
Figure 2. Phenotypic changes during modern wheat breeding in China and the United States. **A)** Summary of phenotypic changes of the 21 agronomic traits investigated. The figure shows phenotypic changes of the 21 traits (upper panel) and a heatmap of the traits (lower panel). The numbers 1 and 2 indicate the significance levels of phenotypic comparisons between CHN_CV vs. Landrace and USA_CV vs. Landrace. The numbers 3 and 4 indicate the differences in phenotypes between CHN_CV vs. Landrace and USA_CV vs. Landrace. Up, Down, and NS indicate an increase, decrease, or no significant difference in pairwise comparisons of the average BLUP values of CHN_CV vs. Landrace and USA_CV vs. Landrace. YPP, yield per plant; HI, harvest index; SR, seed roundness; TNMS, tiller number during the mature stage; AL, awn length; AD, anthesis days; PL, peduncle length; FLL, flag leaf length; PH, plant height; SPL, spike length; TKW, thousand kernel weight; SNPS, seed number per spike; FLW, flag leaf width; SW, seed width; SD, stem diameter; BPP, biomass per plant; HD, heading days; TNSS, tiller number during the seedling stage; SL, seed length; SN, spikelet number; SSN, sterile spikelet number. **B to F)** Boxplots of the BLUP values of PH **B)**, TKW **C)**, SL **D)**, HD **E)**, and TNMS **F)** among subpopulations. The lower and upper lines of each box denote the 25th and 75th percentiles, respectively. The middle lines in the boxes represent medians. The upper whiskers indicate the maximum or 1.5× the interquartile range (IQR). The lower whiskers indicate the minimum or 1.5× the IQR. **P < 0.01; *P < 0.05; NS, no significant difference (for **A to F**). **G to I)** Phenotypes of whole plants **G)**, spikes **H)**, and seeds **I)** represented by PI 321967 for Landrace, Jimai 20 for CHN_CV, and Ripper for USA_CV.
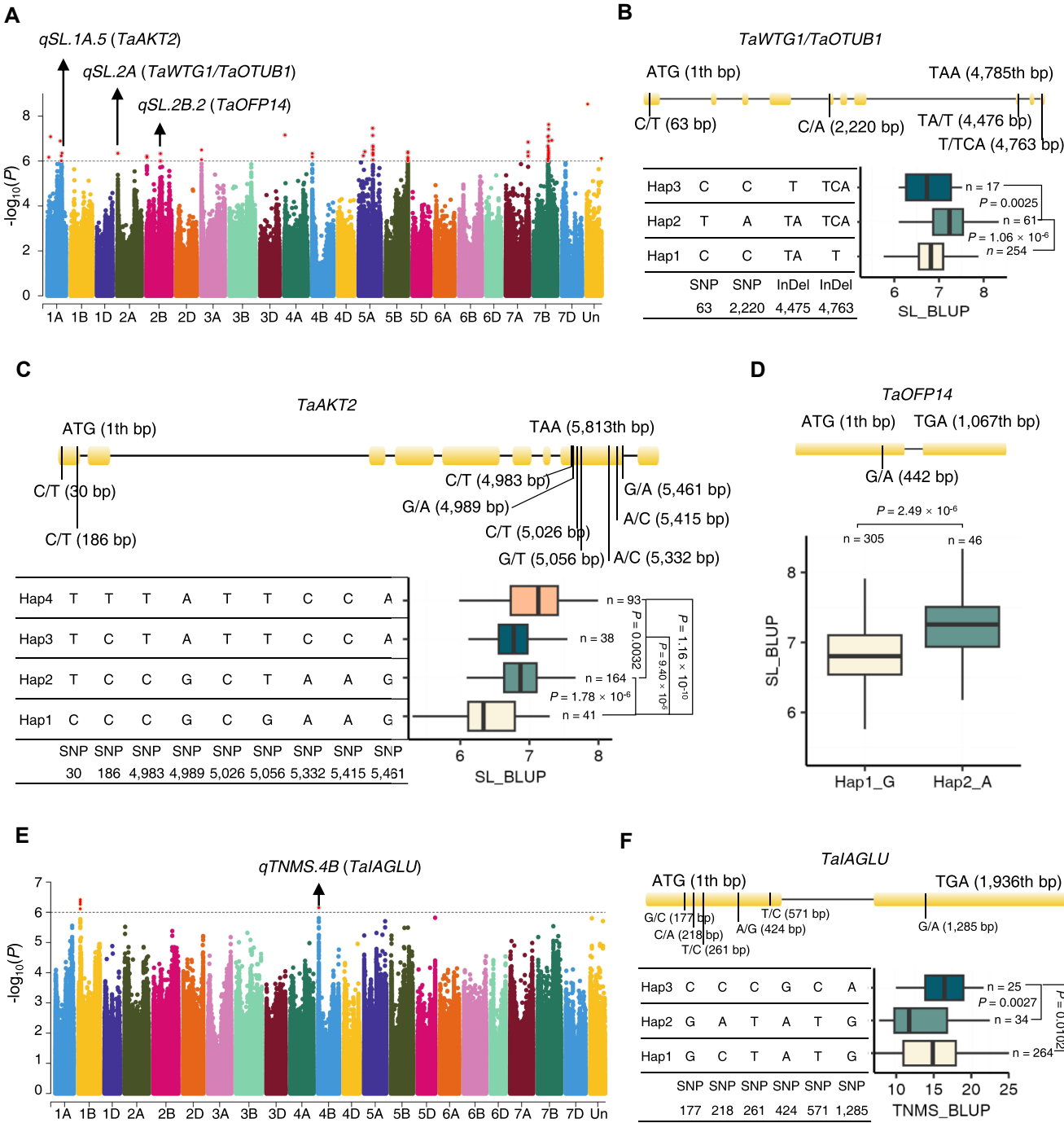
Figure 3. GWAS identification of candidate genes and haplotype analysis. **A)** Manhattan plot of GWAS for SL. The candidate genes are marked above the significantly associated peaks. The dotted line represents the significance threshold ($-\log_{10}(P) = 6$). **B to D)** Candidate gene structure and haplotype analysis of *TaWTG1/TaOTUB1* **B)**, *TaAKT2* **C)**, and *TaOFP14* **D)** based on putative polymorphic variations. The plots contain candidate gene structures and putative variations (upper), haplotypes (lower left), and boxplots of the corresponding phenotypic BLUP values. *n* indicates the number of accessions in each haplotype. *P* indicates the *P*-value based on the 2-tailed *t* test. **E)** Manhattan plot of GWAS for TNMS. **F)** Gene structure and haplotype analysis of the candidate gene *TaIAGLU*.

We also identified genomic regions and candidate genes for other key agronomic traits (Supplemental Data Set 11). These findings will be useful for further functional dissection of trait variation in wheat and its improvement.

Accumulation of favorable alleles in modern wheat cultivars

To investigate the effects of improvement selection on alleles significantly associated with variation in major agronomic traits, we compared the favorable allele frequency (FAF) between the landraces and the cultivars from China and the United States. Favorable alleles were defined as alleles associated with earlier HD and AD; reduced PH, PL, and SSN; and increased SPL, FLL, TKW, seed number per spike (SNPS), SW, tiller number during the seedling stage (TNSS), TNMS, SN, SL, YPP, BPP, HI, flag leaf width (FLW), seed roundness (SR), AL, and stem diameter (SD). We identified 71.98% (149/207) and 69.57% (144/207) of GWAS-associated leading SNPs with increased FAF in the CHN_CV and USA_CV subgroups compared with Landrace, respectively (Fig. 4A; Supplemental Data Set 10). Similar results (63.80%, 3,784/5,931 in CHN_CV; 66.58%, 3,949/5,931 in USA_CV) were obtained by directly analyzing all the GWAS hits ($P < 1e^{-6}$; Fig. 4A; Supplemental Data Set 10). The increase in FAF for most loci suggests that positive selection at specific genomic regions underlies the corresponding agronomic traits.

To assess differences in the direction of selection acting on alleles positively affecting agronomic traits in China and the United States, we investigated the direction of changes in FAF between the cultivars from these 2 regions and the landraces. Among these loci, 67.15% (139/207) showed increased FAF and 25.60% (53/207) showed decreased FAF in both countries (Fig. 4, B and C; Supplemental Data Set 10), suggesting that selection acted in similar directions in China and the United States. Furthermore, 7.03% (15/207) of loci showed different directions of selection between the 2 countries (Fig. 4, B and C; Supplemental Data Set 10), which is suggestive of different targets between the 2 countries. For example, the FAF of lead SNP s_2D_023156697 (around *Rht8*) related to PH increased (by 0.77) in CHN_CV and slightly decreased (by 0.06) in USA_CV, which is consistent with the observation that the dwarf allele of *Rht8* was widely used in the wheat cultivation area around the Yellow and Huai River Valleys of China (Xiong et al. 2022). For each agronomic trait except FLW (33.3%), the proportions of variants showing increased FAF were >76.9% (Fig. 4, D to M; Supplemental Figs. S10 and S11), suggesting that modern wheat breeding in China and the United States favored the same beneficial alleles.

To assess whether wheat improvement was accompanied by the accumulation of favorable alleles at multiple loci across the genome, we analyzed the allelic composition of each of the 355 wheat accessions at 207 trait-associated markers (Fig. 5A; Supplemental Data Set 12). We observed substantial increases in the proportions of favorable alleles at

these loci (Wilcoxon test, $P = 2.20 \times 10^{-16}$ in CHN_CV vs. Landrace and $P = 5.31 \times 10^{-12}$ in USA_CV vs. Landrace) in the modern wheat populations from China and the United States (Fig. 5B). The accumulation of favorable alleles in each accession for all analyzed traits had additive effects (Fig. 5, C and D; Supplemental Fig. S12). The average phenotypic values for HD, AD, PH, PL, and SSN tended to decrease with increasing number of favorable alleles, while SPL, FLL, TKW, SW, TNSS, TNMS, SN, SL, YPP, BPP, HI, FLW, SR, and AL showed the opposite trend. These results suggest that pyramiding favorable alleles identified in this study could facilitate the breeding of high-yielding wheat varieties in the future.

Shared and unique wheat breeding targets in China and the United States

In addition to genetic loci underlying morphological features of wheat, genetic loci providing adaptation to local climatic and environmental conditions could also be targeted by selection during wheat improvement (Wang et al. 2020). We used the cross-population composite likelihood ratio (XP-CLR) test (Chen et al. 2010) to identify selection signatures associated with wheat improvement in China and the United States by comparing CHN_CV vs. Landrace and USA_CV vs. Landrace populations. Using the top 5% of XP-CLR scores as a threshold, we detected 2,037 and 1,866 selective sweeps in CHN_CV and USA_CV, respectively, which covered 2,341.51 Mb (16.10%) and 2,198.84 Mb (15.11%) of the common wheat genome (IWGSC RefSeq v1.0; Fig. 6; Supplemental Fig. S13A and Data Sets 13 and 14) (International Wheat Genome Sequencing Consortium (IWGSC) 2018). These selected genomic regions show reduced nucleotide diversity compared with the rest of the genome, and the divergence between the cultivars and landraces in these genomic regions has increased (2-tailed *t* test, $P = 9.66 \times 10^{-41}$; Supplemental Data Set 15). The selected regions are distributed unevenly along the wheat chromosomes (Supplemental Fig. S13A), suggesting asymmetric selection among the 3 wheat subgenomes. The selected genomic regions encompass 18,241 and 16,692 genes in CHN_CV and USA_CV, respectively (Supplemental Data Sets 13 and 14). Among these, 5,428 genes overlapped with the regions selected in both CHN_CV and USA_CV, indicating shared breeding targets between the 2 countries (Supplemental Fig. S13B and Data Sets 13 and 14).

Genes with experimentally verified functions (known genes) related to plant architecture, growth period, starch synthesis, resistance to foliar diseases, and end-use quality were also located in the genomic regions of selective sweeps (Supplemental Data Sets 16 and 17). *Rht-A1* and *TaERF8-2B*, which are associated with PH and grain yield, respectively, were located in selective sweeps detected in both populations of cultivars (Fig. 6, A and B; Supplemental Data Sets 16 and 17) (Peng et al. 1999; Zhang et al. 2020). *Rht-D1* (*Rht2*), *Rht18*, and *TaGA2ox8* showed evidence of selection
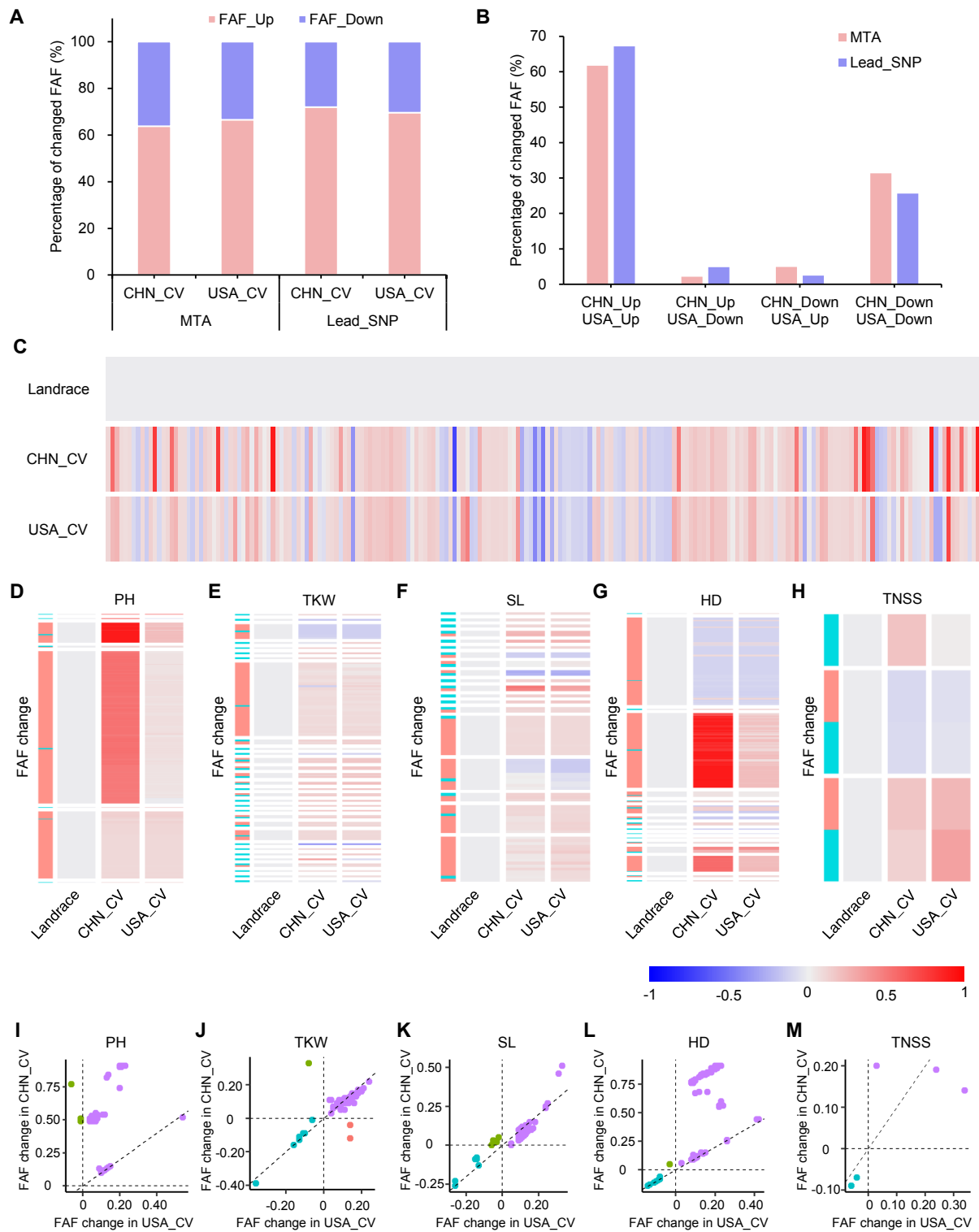
Figure 4. Changes in FAF during the modern wheat breeding process. **A)** Percentages of lead SNPs and MTAs with increased and decreased FAF. FAF_Up and FAF_Down indicate SNPs with increased or decreased FAF in CHN_CV vs. Landrace and USA_CV vs. Landrace. **B)** The percentages of lead SNPs and MTAs with simultaneously increased or decreased FAF. **C)** Heatmap of changes in FAF for 207 lead SNPs in each subpopulation. Each column stands for a lead SNP. **D to H)** Heatmap of FAF changes for MTAs ($P < 1e^{-6}$) for PH **D)**, TKW **E)**, SL **F)**, HD **G)**, and TNSS **H)** in the subpopulations. Red represents an increase in FAF, and blue represents a decrease. Each row represents an associated SNP. Cyan and pink in the first column indicate the lead SNPs and the remaining MTAs, respectively ($P < 1e^{-6}$). White blanks in rows are used to separate each locus (containing lead SNP and MTAs). **I to M)** Dot plots of FAF changes for MTAs of PH **I)**, TKW **J)**, SL **K)**, HD **L)**, and TNSS **M)**. Purple, light green, blue, and light red indicate points in the first, second, third, and fourth quadrants, respectively. The dashed lines represent $x = 0$, $y = 0$, and $y = x$.
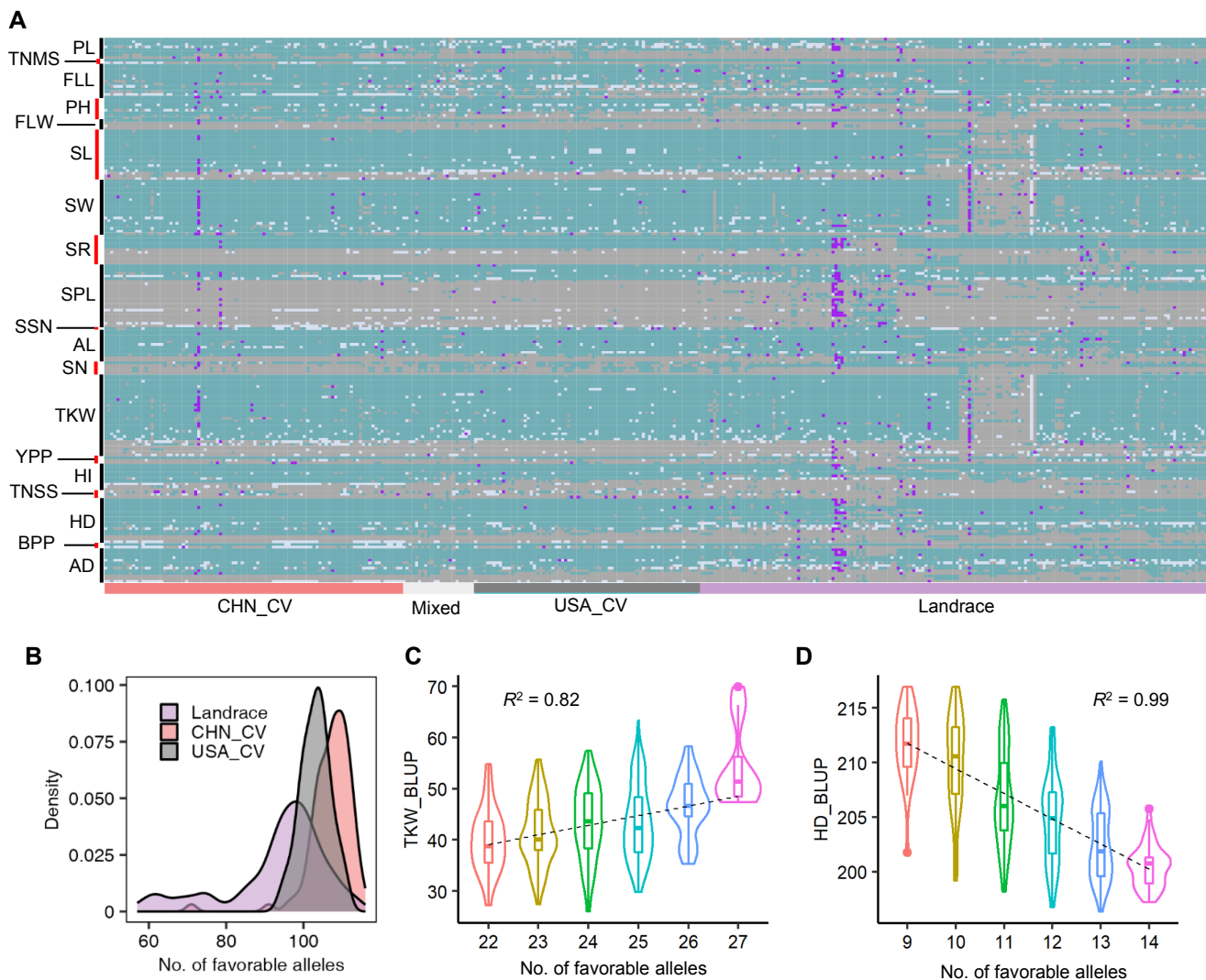
Figure 5. Genomic fingerprinting analysis of the 355 common wheat accessions for lead SNPs. **A)** The genomic fingerprints of the 355 common wheat accessions for 207 lead SNPs associated with key traits (each row indicates a lead SNP, and each column indicates an accession). Lead SNPs associated with the same trait are grouped together, and the trait names are marked to the left of the plot. Dark green, gray, purple, and light white indicate favorable, undesirable, heterozygous, and missing alleles, respectively. **B)** Density distribution of favorable alleles in each common wheat accession. **C, D)** Relationships between the number of favorable alleles and phenotypic values for TKW **C)** and HD **D)**. The violin plots and boxplots were combined to display the distribution of the average values of phenotypic traits. The dashed lines represent linear relationships between the phenotypic values and number of favorable alleles, which is fitted by $y \sim x$. R^2 indicates the correlation coefficient of the phenotypic data and the number of favorable alleles.

only in China, whereas the *Rht-B1* (*Rht1*) locus was uniquely selected in the United States (Peng et al. 1999; Ford et al. 2018; Sun et al. 2018). Six known genes (*Vrn2, TaFT3-1, TaFT4-1, TaGRP-2, TaVRT-2*, and *TaAGL12*) controlling growth and development were found in the shared selective sweep regions detected in both countries (Yan et al. 2004; Shimizu et al. 2020). We also identified known genes related to growth period under country-specific selection. For example, the photoperiod gene *PPD1* and the vernalization gene *Vrn3* were selected only in China, while the MADS-box gene *TaSEP3-A1* was targeted in the United States (Fig. 6A; Supplemental Data Sets 16 and 17) (Yan et al. 2006; Nishida et al. 2013;

Zhang et al. 2021). Furthermore, the selective sweep regions containing known genes involved in starch synthesis, grain yield, grain size, and flour quality are considered to have been under strong selection during modern wheat breeding for higher yield and end-use quality. Six known genes related to starch synthesis, grain yield, or flour quality (*Glu-B3, TaGBSSII-2A, TaSSIIIb-2D, TaAGPS1-a, TaCwi-5D*, and *TaGW8-7B*) are present in shared selective sweep regions. Twenty-two and 21 known genes associated with starch synthesis, grain size, or grain yield traits have been selected in China and the United States, respectively (Fig. 6A; Supplemental Data Sets 16 and 17).
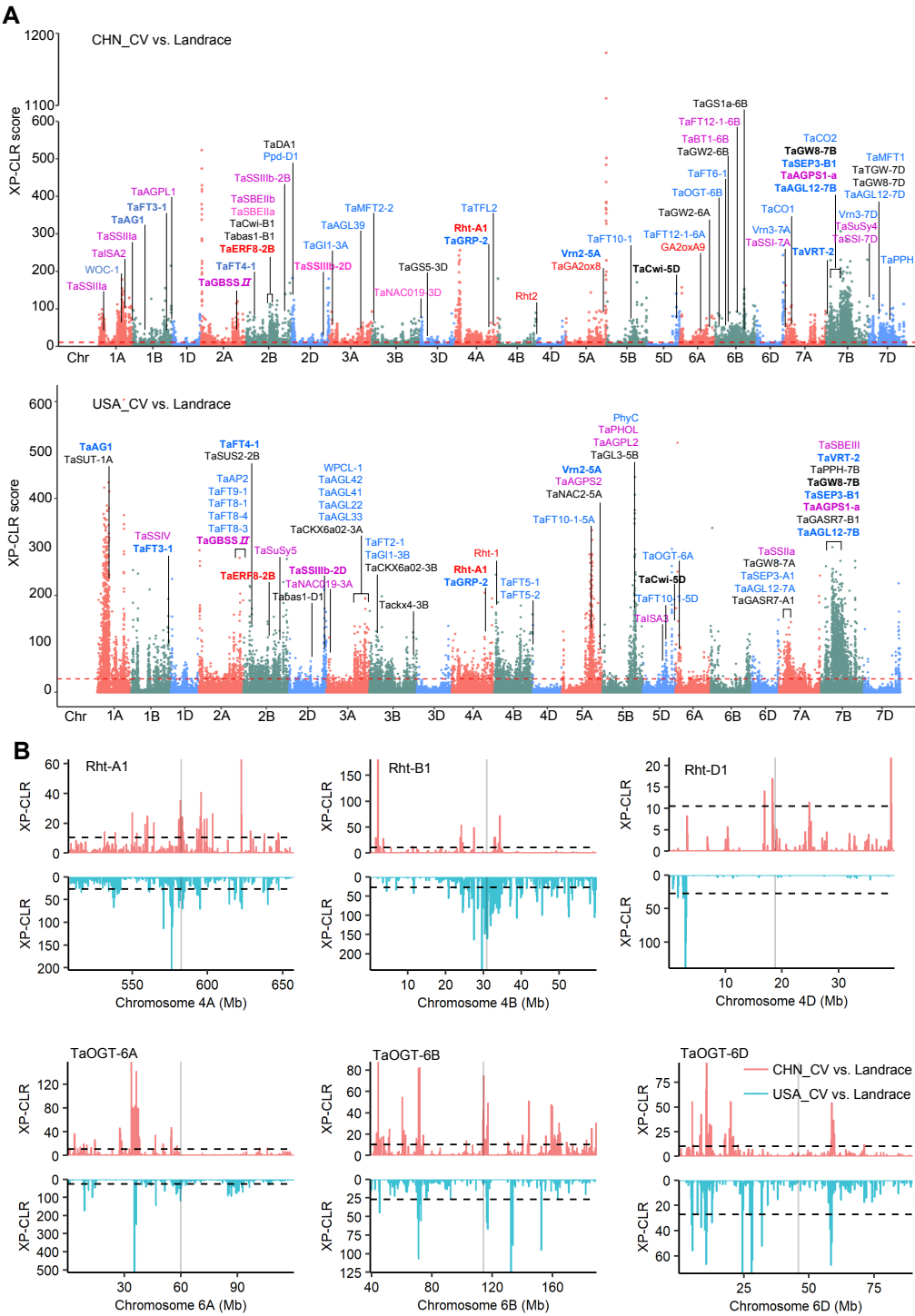
Figure 6. Profiling of selective sweeps during modern wheat breeding. **A)** Genome-wide selective sweeps in CHN_CV vs. Landrace (upper) and USA_CV vs. Landrace (lower). The red horizontal dashed lines indicate the cutoffs of the top 5% of values. Known genes that overlap with selective sweeps are marked above the selection signals. Known genes related to PH, growth period, starch synthesis, and grain yield are indicated in red, blue, purple, and black, respectively. Known genes in bold were potentially selected in both CHN_CV and USA_CV. **B)** Comparison of XP-CLR score selection signals for *Rht* and *TaOGT* among the A, B, and D subgenomes. The gray vertical lines indicate the positions of the corresponding genes. The black horizontal dashed lines indicate the cutoffs of the top 5% of values.

Interestingly, some genes were under potential selection in only 1 country, whereas their homologs were selected in the other country. For example, the *O*-linked *N*-acetylglucosamine transferase gene *TaOGT-6B* (Fan et al. 2021), which regulates flowering time, was selected in China, whereas its homolog in the A subgenome (*TaOGT-6A*) was selected in the United States (Fig. 6B). *TaGW8-7B*, a gene related to grain size and TKW, is located in selective sweep regions detected in both countries. However, its homolog in the D subgenome (*TaGW8-7D*) was selected in China, whereas its A subgenome homolog (*TaGW8-7A*) was selected in the United States (Supplemental Fig. S14A) (Yan et al. 2019). Another set of 8 known homoeologous gene triplets showed the same mode of country-specific selection (Supplemental Data Sets 16 and 17). These results suggest that improvement selection targeted different homoeoalleles in different countries to achieve common breeding goals.

Of the 207 loci detected in GWAS for the agronomic traits examined in our study, 90 and 59 are colocated with the selective sweep regions detected in CHN_CV and USA_CV, respectively. Among these, 36 loci are shared between China and the United States, implying that these loci were subjected to convergent selection in the 2 countries. Finally, 54 and 23 selected loci were unique to China and the United States, respectively (Supplemental Fig. S14B and Data Sets 18 and 19). These results further demonstrate that breeding efforts in China and the United States targeted both shared and distinct genomic regions.

Discussion

Frequent extreme weather events, the growing population, and the genetic erosion of crops (Khoury et al. 2022) are putting huge pressure on the global food supply. However, conventional breeding approaches capable of increasing grain yield by approximately 1% annually for most crops, including wheat, cannot satisfy the rising food demand (Ray et al. 2013). Genetic improvement has proven to be successful for generating crops with higher yields, stronger adaptation, and better human end-use properties (Xie et al. 2015). Nevertheless, diverse strategies have likely been implemented in different geographical regions during crop improvement given the differences in farming systems, human preferences, and local climates. Therefore, investigating their genetic basis may pave the way for further advances in crop breeding.

In this study, we developed a whole-genome diversity map for a panel of common wheat landraces and cultivars collected from China and the United States and applied comparative population genomics and association genetics to investigate the impact of breeding on the plant architecture and genetic composition of regional wheat populations. We showed that plant architecture has been targeted convergently, but with different magnitudes, in regional breeding programs to achieve the common breeding goals of higher yields, better quality, and wider adaptability. Considering the diverse origins of the materials examined in this study,

some factors affecting local adaptation might not have been fully expressed under our experimental conditions. Consistent with the significant phenotypic changes during wheat improvement, wheat genetic diversity decreased approximately by 20% (18.9% in China and 29.1% in the United States), which is higher than the previously reported value (~5%) (Cavanagh et al. 2013). Perhaps the ascertainment bias associated with the inclusion of common variants into the SNP genotyping array used in that study did not allow the majority of the rare variants in the populations to be captured.

We identified 207 loci underlying key agronomic traits by performing GWAS with tens of millions of markers in a panel of diverse germplasm. The beneficial allele frequency of most agronomic traits increased during modern wheat breeding in both countries, suggesting convergent positive selection at these loci. Despite the substantial differences in the genetic compositions of the cultivars from China and the United States, the same direction of phenotypic changes was observed for 8 of the 21 agronomic traits investigated. This finding is consistent with the direct or indirect contributions of the traits analyzed in this study to grain yield, which is the main target of breeding efforts.

Furthermore, we identified a number of candidate selective signatures representing potential breeding targets. These genomic regions contain genes involved in regulating plant architecture, starch synthesis, heading date, and grain yield and are partially shared between China and the United States. The partial sharing of candidate genes under selection is likely due to a series of factors, such as diverse environmental conditions, linkage drag, random genetic drift, different selection pressures, and multiple functionally equivalent mutations. The genomic regions under unique selection in either China or the United States could serve as critical resources for incorporating successfully adapted alleles into future wheat improvement efforts.

We also identified genes (for example: *TaSSIIIb-2D* encoding a starch synthesis enzyme; Supplemental Data Sets 16 and 17 and Fig. S15) that were convergently selected in China and the United States by targeting totally different haplotypes, suggesting that convergent selection in the 2 countries was achieved by targeting the same genes with different types of variants. Our results also highlight the role of polyploidy in the evolution of agronomic traits in wheat, as manifested by country-specific selection acting on the allelic variants of homoeologous genes from different subgenomes. The existence of region-specific targets of improvement selection suggests that an underutilized genetic diversity is available for wheat improvement in each of the 2 analyzed populations of cultivars from China and the United States.

In conclusion, we generated a comprehensive landscape of genomic variation in a diverse panel of wheat cultivars and landraces. We also constructed a genome-to-phenome association map and an atlas of selective sweeps linked to wheat improvement, representing valuable resources for future functional genetic studies and wheat improvement.

Materials and methods

Wheat accessions and phenotypic measurements

In total, we collected 355 common wheat (*T. aestivum*) accessions, including 180 landraces, 103 cultivars from China, and 72 cultivars from the United States. The 180 landraces originated from 13 countries around the Fertile Crescent and were selected from a worldwide population of landraces from the United States Department of Agriculture (USDA) to represent a pool of genetically diverse germplasms. The representativeness of the 180 landraces was evaluated by comparing them with 632 common wheat accessions, which were collected from 69 countries and described as a worldwide landrace population (Balfourier et al. 2019). The 180 landraces selected in this study revealed good representativeness and harbored more than 94% of the genetic diversity of the worldwide wheat landrace collection (Supplemental Fig. S1). The genetic diversity (π) was evaluated by identifying the common SNPs shared between the 180 and 632 landraces. We also evaluated the representativeness of the 180 landraces by examining the genetic distances of pairwise of accessions between the 180 landraces and 632 published landraces using the method reported by Schulthess et al. (2022). We calculated the identity by state (IBS) using the snpgdsIBSNum function of SNPRelate and the proportion of pairwise difference (PPD) between 2 samples using the following formula: IBS0/(IBS0 + IBS2). PPD values were used to evaluate the representativeness between 2 subpopulations (180 and 632 landraces). An accession was considered to have counterparts in the other subpopulation when its minimum genetic distance to all accessions of the other subpopulation was less than the 95% quantile of the distances within the 632-member collection. A set of 103 and 72 cultivars developed after the Green Revolution were collected privately from breeding programs in China and the United States, respectively (Supplemental Data Set 1).

All 355 common wheat accessions were planted (winter growing seasons) in 3 consecutive years (2013 to 2016) in Zhao County, Shijiazhuang City, Hebei province, China (38° 05′N, 114°52′E). The accessions were randomly arranged in plots with a row and column spacing of 110 cm × 25 cm and 3 independent replicates. Five plants in the middle of each plot were selected to evaluate the agronomic performance of each accession. YPP represents the mean seed weight of 5 individual plants. TKW was calculated by dividing the YPP by the number of seeds per plant and then multiplying by 1,000. The number of seeds per plant, SL, SW, and SR were measured using a Crop Grain Appearance Quality Scanning Machine (SC-E, Wanshen Technology Company, Hangzhou, China). BPP is the average weight of 5 whole plants (without roots) during the mature stage, and HI was obtained by dividing YPP by BPP. HD and AD were calculated as the days from sowing to heading/anthesis of half of the spikes in a row at the spike emerging stage. TNSS and TNMS are the average of 5 plants during Zadok 29 (plants containing a main stem and 9 or more tillers) and mature stage,

respectively. At maturity, PH (aboveground and excluding awns), PL, AL, SPL, SN, SSN, and SNPS (main spikes) were obtained by averaging the values from 5 plants. After anthesis, FLL was measured as the distance from leaf bottom to leaf tip. FLW is the length of the widest part of the flag leaf. SD is the diameter of the peduncle stem (2 cm above the stem joint).

DNA isolation and sequencing

Young leaf tissue from each accession was sampled to extract genomic DNA using the cetyltrimethylammonium bromide (CTAB) method. PCR-free DNA libraries with an insert size of 350 bp were constructed and sequenced with the DNBSEQ platform at BGI-Shenzhen, yielding a total of $\sim 8.0 \times 10^{11}$ 100-bp paired-end reads and an average depth of coverage at 14.52× for each accession.

Variant calling, quality control, and annotation of genetic variants

Raw reads were trimmed with Trimmomatic (version 0.36) to control read quality (Bolger et al. 2014). The clean reads were then mapped to the Chinese Spring wheat reference genome (IWGSC RefSeq v1.0) using the "mem" module in BWA with default parameters (Li and Durbin 2009). A genomic variant call format (GVCF) file for each sample was obtained using HaplotypeCaller in Genome Analysis Toolkit (GATK, Version 3.7-0-gcfedb67) (McKenna et al. 2010). All GVCF files were then used for joint genotyping to obtain a single VCF file for all wheat lines. Given the bias caused by misalignment, we discarded variant sites with too many (average coverage depth more than 30×) or too few (less than 3×) total aligned reads (DP < 1,700 or DP > 10,400). Variant quality was controlled based on the following criteria: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || ReadPosRankSum < −8.0 || MQRankSum < −12.5" for SNPs and "QD < 2.0 || FS > 200.0 || ReadPosRandSum < −20.0" for InDels. To further control variant quality, we removed variant sites with "QUAL < 150." At the population level, we discarded variants with a missing rate > 25% and a heterozygosity > 30%. For InDels, we considered only InDels ≤ 8 bp. The MAF was set to 0.01 for statistical analysis of variants (76,874,471 SNPs and 5,208,800 InDels), including genetic diversity and genetic differentiation. SNPs with MAF > 0.05 (44,050,985) were used for GWAS and other analyses.

To evaluate the accuracy of the SNPs identified in this study, we compared the genotypes of SNP sites shared between the resequencing data and the wheat 660K SNP array data (http://wheatomics.sdau.edu.cn/download.html) for 343 of the 355 accessions. The flanking sequences of SNPs on the wheat 660K array were aligned to IWGSC RefSeq v1.0. Sequences with a maximum of 1 mismatch, without gaps, and aligned to only 1 position were retained for further analysis. Shared SNPs between the resequencing and SNP array data were used to evaluate the genotyping accuracy of each accession.

To evaluate the variant effects (including SNPs and InDels), we annotated them with SnpEff using HC gene models in the RefSeq v1.0 genome assembly (International Wheat Genome Sequencing Consortium (IWGSC) 2018). Variants with a "HIGH" effect were regarded as deleterious, including start codon loss/gain, stop codon loss/gain, and splice acceptor/donor variants.

Statistical analysis of phenotypes

To calculate the best linear unbiased prediction (BLUP) values for 21 agronomic traits, we fitted the phenotypic values into the following formula: $Y_{ij} = \mu + \text{Line}_i + \text{Year}_j + (\text{Line} \times \text{Year})_{ij} + (\text{Year} \times \text{Rep})_{jn} + \text{error}_{ijn}$ (where μ represents the mean of the phenotypic values, Line_i is the genotype effect of the ith accession, Year_j is the effect of the jth year, $(\text{Line} \times \text{Year})_{ij}$ and $(\text{Year} \times \text{Rep})_{jn}$ are the effects of genotype–year and year–replication interactions, respectively, and error_{ijn} is the error of the random effect), using a mixed linear model in R with the lme4 package (Bates et al. 2014). Phenotypic comparisons between subpopulations were performed with ggsignif (version 0.6.3) and ggplot2 (Ahlmann-Eltze and Patil 2021).

PCA

Considering the large LD distance and high-density SNP distribution, we obtained a random subset of relatively independent SNPs using PLINK1.9 (Purcell et al. 2007) with the following criteria: (i) the LD coefficient (r^2) was less than 0.4 for any pair of SNPs in a window of 1,000 consecutive SNPs with a step size of 10 SNPs (–indep-pairwise 1000 10 0.4), and (ii) the SNPs on unanchored scaffolds (chrUn) were excluded from the subset. After pruning, a subset of 304,744 SNPs was obtained for PCA using PLINK1.9.

Population genetic diversity, differentiation, and structure

The genetic diversity (π) in each subpopulation (Landrace, CHN_CV, and USA_CV) and the genetic differentiation (F_{ST}) between subpopulations were calculated using a 500-kb sliding window and a step size of 100 kb with VCFtools (v0.1.16) (Danecek et al. 2011). The minimum F_{ST} values were set to 0 when we calculated the mean F_{ST} values.

Genetic assignment analysis was conducted using the pruned subset of 304,744 SNPs with the ADMIXTURE program (Alexander et al. 2009). A total of 10 independent runs of ADMIXTURE with different random seeds were performed at each K and aligned with CLUMPP (Jakobsson and Rosenberg 2007). f_3 statistics (Reich et al. 2009) were calculated to examine admixture or copopulation membership among all 3-way triplets between different subpopulations using TreeMix (Pickrell and Pritchard 2012).

LD analysis

To estimate and analyze the LD decay patterns for all samples and different populations, we randomly selected 1% of all SNPs using the parameter "--thin 0.01" in PLINK. We then calculated the squared correlation coefficient (r^2) between pairwise SNPs using PopLDdecay with the parameter "-MaxDist 10000" (Zhang et al. 2019).

GWAS and candidate gene prediction

A GWAS was performed with a linear mixed model that accounted for both population structure and kinship for all 21 agronomic traits using the "mlma" module with the "--mlma" parameter in GCTA (Yang et al. 2011). The first 3 principal components mentioned above were used to control the population structure. A kinship matrix accounting for pedigree relationships was calculated from the subset of independent SNPs using GCTA. A threshold of 1×10^{-6} (Benjamini–Hochberg FDR < 0.05) was used to identify significantly associated SNPs, which were then delineated into QTLs based on physical distance. Significantly associated SNPs with a physical distance > 5 Mb were regarded as independent QTLs. The SNP with the lowest P-value within a QTL was designated as the lead SNP to represent the corresponding QTL. The 1-Mb genomic regions centered on the lead SNPs were used to identify candidate genes. Based on the VCF files annotated by SnpEff, missense or frameshift mutations in the promising candidate genes were inferred to be the causal polymorphisms. Haplotype analyses were performed based on the causal polymorphisms of the corresponding candidate genes. Haplotypes in fewer than 10 accessions were discarded. Phenotypic comparisons between 2 haplotypes were performed by 2-tailed t tests. The coding sequences of candidate genes with different haplotypes were sequenced and compared with published reference sequences to avoid larger InDels. To obtain functional annotations for genes in each candidate region, the encoded protein sequences of every gene were searched against the rice protein database (MSU Version 7.0) by BLASTp. An E-value threshold of 1×10^{-5} was used to identify rice homologs of the candidate genes.

Frequency and genomic fingerprints of agronomically favorable alleles

Based on the GWAS results, we analyzed the FAF of marker–trait-associated SNPs (MTAs, $P < 1e^{-6}$) and lead SNPs in each subpopulation. The alleles with earlier HD and AD, reduced PH, PL, and SSN, or increased SPL, FLL, TKW, SNPS, SW, TNSS, TNMS, SN, SL, YPP, BPP, HI, FLW, SR, AL, and SD were designated as favorable alleles. Changes in FAF between CHN_CV vs. Landrace and USA_CV vs. Landrace were used for further analysis. To construct genomic fingerprints, we extracted a subset of VCF files for leading SNPs. We used the numbers 0, 0.5, and 1 to represent undesirable, heterozygous, and favorable alleles at each locus, respectively (Supplemental Data Set 12).

Detection of selective sweeps between landraces and cultivars

We used a python version of the composite likelihood approach (XP-CLR) (Chen et al. 2010) to identify selective

sweeps during modern wheat improvement (https://github.com/hardingnj/xpclr). Landraces were regarded as the reference, and CHN_CV and USA_CV were used as the queries. We scanned for selective sweeps with a step size of 100 kb and a 500-kb sliding window across each chromosome (−size 500,000; --step 100,000). The maximum number of SNPs allowed in each window was set to 500 using the parameter "--maxsnps 500." The genomic regions with XP-CLR scores above the 95th percentile were considered to be under selection. Adjacent selective sweep windows < 300 kb apart were merged. We regarded HC gene models in the IWGSC RefSeq v1.1 annotation in the selective sweep regions as potential breeding targets. The lead SNPs for each GWAS hit closest to selective sweeps (<500 kb) were regarded as overlapping hits.

Statistical analysis

Statistical analyses were performed as described in each figure legend. Statistical data are provided in Supplemental Data Set 20.

Accession numbers

Accession numbers of the genes discussed in the main text are provided in Supplemental Data Set 21.

Author contributions

H.-Q.L. and Shu.Z. conceived the project. J.N. conducted the bioinformatic and statistical analyses and drafted the manuscript. S.M. assisted in data analyses and discussing the results. H.-Q.L., Shu.Z., and F.H. supervised the project. Shu.Z. and Y.L. were responsible for field trials and collected phenotypic data. Shan.Z., C.Z., and Y.H. were responsible for sequence data generation and variant calling. J.P. provided some of the germplasms. Y.S., S.T., X.S., H.S., H.W., Y.C., M.K.N., C.C., W.L., Y.Z., M.G., M.C., Q.L., X.L., and W.Z. helped with field trials. E.A. edited and revised the manuscript. H.-Q.L., Shu.Z., F.H., and S.M. revised the manuscript.

Supplemental data

The following materials are available in the online version of this article.

Funding

Data availability

The raw data and sequencing data have been deposited in the National Genomics Data Center (NGDC) database under project CRA005878. The genotype data in variant call formats (VCF) are available in the NGDC database under accession number GVM000315.

References

Ahlmann-Eltze C, Patil I. ggsignif: R package for displaying significance brackets for 'ggplot2'. PsyArxiv. 2021.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009:**19**(9): 1655–1664. https://doi.org/10.1101/gr.094052.109

Balfourier F, Bouchet S, Robert S, De Oliveira R, Rimbert H, Kitt J, Choulet F; International Wheat Genome Sequencing Consortium; BreedWheat Consortium; **Paux E.** Worldwide phylogeography and history of wheat genetic diversity. Sci Adv. 2019:**5**(5):eaav0536. https://doi.org/10.1126/sciadv.aav0536

Bates D, Mchler M, Bolker B, Walker SJJoss. Fitting linear mixed-effects models using lme4 arXiv:1406. 2014.

Bhatta M, Morgounov A, Belamkar V, Poland J, Baenziger PS. Unlocking the novel genetic diversity and population structure of synthetic Hexaploid wheat. BMC Genomics 2018:**19**(1):591. https:// doi.org/10.1186/s12864-018-4969-2

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014:**30**(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci U S A. 2013:**110**(20):8057–8062. https://doi.org/10.1073/pnas.1217133110

Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010:**20**(3):393–402. https://doi.org/10.1101/gr.100545.109

Cheng H, Liu J, Wen J, Nie X, Xu L, Chen N, Li Z, Wang Q, Zheng Z, Li M, et al. Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. Genome Biol. 2019:**20**(1):136. https://doi.org/10.1186/s13059-019-1744-x

Choi MS, Koh EB, Woo MO, Piao R, Oh C-S, Koh H-J. Tiller formation in rice is altered by overexpression of *OsIAGLU* gene encoding an IAA-conjugating enzyme or exogenous treatment of free IAA. J Plant Biol. 2013:**55**(6):429–435. https://doi.org/10.1007/s12374-012-0238-0

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics 2011:**27**(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DeWitt N, Guedira M, Lauer E, Sarinelli M, Tyagi P, Fu D, Hao Q, Murphy JP, Marshall D, Akhunova A, et al. Sequence-based mapping identifies a candidate transcription repressor underlying awn suppression at the *B1* locus in wheat. New Phytol. 2019:**225**(1): 326–339. https://doi.org/10.1111/nph.16152

Du D, Zhang D, Yuan J, Feng M, Li Z, Wang Z, Zhang Z, Li X, Ke W, Li R, et al. *FRIZZY PANICLE* defines a regulatory hub for simultaneously controlling spikelet formation and awn elongation in bread wheat. New Phytol. 2021:**231**(2):814–833. https://doi.org/10.1111/nph.17388

Fan M, Miao F, Jia H, Li G, Powers C, Nagarajan R, Alderman PD, Carver BF, Ma Z, Yan L. *O*-Linked *N*-acetylglucosamine transferase is involved in fine regulation of flowering time in winter wheat. Nat Commun. 2021:**12**(1):2303. https://doi.org/10.1038/s41467-021-22564-8

Fang L, Wang Q, Hu Y, Jia Y, Chen J, Liu B, Zhang Z, Guan X, Chen S, Zhou B, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. Nat Genet. 2017:**49**(7):1089–1098. https://doi.org/10.1038/ng.3887

Ford BA, Foo E, Sharwood R, Karafiatova M, Vrána J, MacMillan C, Nichols DS, Steuernagel B, Uauy C, Doležel J, et al. *Rht18* semi-dwarfism in wheat is due to increased *GA 2-oxidaseA9* expression and reduced GA content. Plant Physiol. 2018:**177**(1):168–180. https://doi.org/10.1104/pp.18.00023

Gaire R, Ohm H, Brown-Guedira G, Mohammadi M. Identification of regions under selection and loci controlling agronomic traits in a soft red winter wheat population. Plant Genome 2020:**13**(2):e20031. https://doi.org/10.1002/tpg2.20031

Guo W, Xin M, Wang Z, Yao Y, Hu Z, Song W, Yu K, Chen Y, Wang X, Guan P, et al. Origin and adaptation to high altitude of Tibetan semi-wild wheat. Nat Commun. 2020:**11**(1):5085. https://doi.org/10.1038/s41467-020-18738-5

Hao C, Jiao C, Hou J, Li T, Liu H, Wang Y, Zheng J, Liu H, Bi Z, Xu F, et al. Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. Mol Plant. 2020:**13**(12):1733–1751. https://doi.org/10.1016/j.molp.2020.09.001

He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, Forrest K, Fritz A, Hucl P, Wiebe K, et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nat Genet. 2019:**51**(5):896–904. https://doi.org/10.1038/s41588-019-0382-2

Hu L, Tu B, Yang W, Yuan H, Li J, Guo L, Zheng L, Chen W, Zhu X, Wang Y, et al. Mitochondria-associated pyruvate kinase complexes regulate grain filling in rice. Plant Physiol. 2020:**183**(3):1073–1087. https://doi.org/10.1104/pp.20.00279

Huang K, Wang D, Duan P, Zhang B, Xu R, Li N, Li Y. *WIDE AND THICK GRAIN 1*, which encodes an otubain-like protease with deubiquitination activity, influences grain size and shape in rice. Plant J. 2017:**91**(5):849–860. https://doi.org/10.1111/tpj.13613

Huang D, Zheng Q, Melchkart T, Bekkaoui Y, Konkin DJF, Kagale S, Martucci M, You FM, Clarke M, Adamski NM, et al. Dominant inhibition of awn development by a putative zinc-finger transcriptional repressor expressed at the *B1* locus in wheat. New Phytol. 2019:**225**(1):340–355. https://doi.org/10.1111/nph.16154

International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 2018:**361**(6403):eaar7191. https://doi.org/10.1126/science.aar7191

Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 2007:**23**(14): 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jordan KW, Wang S, Lun Y, Gardiner LJ, MacLachlan R, Hucl P, Wiebe K, Wong D, Forrest KL; IWGS Consortium, **et al.** A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biol. 2015:**16**(1):48. https://doi.org/10.1186/s13059-015-0606-4

Khoury CK, Brush S, Costich DE, Curry HA, de Haan S, Engels JMM, Guarino L, Hoban S, Mercer KL, Miller AJ, et al. Crop genetic

erosion: understanding and responding to loss of crop diversity. New Phytol. 2022:**233**(1):84–118. https://doi.org/10.1111/nph.17733

Kuzay S, Xu Y, Zhang J, Katz A, Pearce S, Su Z, Fraser M, Anderson JA, Brown-Guedira G, DeWitt N, et al. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. Theor Appl Genet. 2019:**132**(9):2689–2705. https://doi.org/10.1007/s00122-019-03382-5

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009:**25**(14): 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li A, Hao C, Wang Z, Geng S, Jia M, Wang F, Han X, Kong X, Yin L, Tao S, et al. Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. Mol Plant. 2022:**15**(3):504–519. https://doi.org/10.1016/j.molp.2022.01.004

Liu J, Rasheed A, He Z, Imtiaz M, Arif A, Mahmood T, Ghafoor A, Siddiqui SU, Ilyas MK, Wen W, et al. Genome-wide variation patterns between landraces and cultivars uncover divergent selection during modern wheat breeding. Theor Appl Genet. 2019:**132**(9): 2509–2523. https://doi.org/10.1007/s00122-019-03367-4

Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K, Aktas H, Ozer E, Ozdemir F, Manickavelu A, et al. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. J Exp Bot. 2015:**66**(12):3477–3486. https://doi.org/10.1093/jxb/erv122

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M; International Wheat Genome Sequencing Consortium; Jakobsen KS, Wulff BB, Steuernagel B, Mayer KF, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science 2014:**345**(6194):1250092. https://doi.org/10.1126/science.1250092

McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G, et al. Agriculture: feeding the future. Nature 2013:**499**(7456):23–24. https://doi.org/10.1038/499023a

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010:**20**(9): 1297–1303. https://doi.org/10.1101/gr.107524.110

Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpffer H, Basterrechea M, König P, et al. Genebank genomics highlights the diversity of a global barley collection. Nat Genet. 2019:**51**(2):319–326. https://doi.org/10.1038/s41588-018-0266-x

Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: advances and applications. Nat Rev Genet. 2011:**13**(2):85–96. https://doi.org/10.1038/nrg3097

Muqaddasi QH, Brassac J, Koppolu R, Plieske J, Ganal MW, Roder MS. *TaAPO-A1*, an ortholog of rice *ABERRANT PANICLE ORGANIZATION 1*, is associated with total spikelet number per spike in elite European hexaploid winter wheat (*Triticum aestivum* L.) varieties. Sci Rep. 2019:**9**(1):13853. https://doi.org/10.1038/s41598-019-50331-9

Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, Laurie DA, Kato K. Structural variation in the 5′ upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1*a identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time. Mol Breed. 2013:**31**(1):27–37. https://doi.org/10.1007/s11032-012-9765-0

Niu J, Zheng S, Shi X, Si Y, Tian S, He Y, Ling H-Q. Fine mapping and characterization of the awn inhibitor *B1* locus in common wheat (*Triticum aestivum* L.). Crop J. 2020:**8**(4):613–622. https://doi.org/10.1016/j.cj.2019.12.005

Pang Y, Liu C, Wang D, St Amand P, Bernardo A, Li W, He F, Li L, Wang L, Yuan X, et al. High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol Plant. 2020:**13**(9):1311–1327. https://doi.org/10.1016/j.molp.2020.07.008

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, et al. 'Green revolution' genes encode mutant gibberellin response modulators. Nature 1999:**400**(6741):256–261. https://doi.org/10.1038/22307

Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012:**8**(11): e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pont C, Leroy T, Seidel M, Tondelli A, Duchemin W, Armisen D, Lang D, Bustos-Korts D, Goué N, Balfourier F, et al. Tracing the ancestry of modern bread wheats. Nat Genet. 2019:**51**(5):905–911. https://doi.org/10.1038/s41588-019-0393-z

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007:**81**(3):559–575. https://doi.org/10.1086/519795

Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. PLoS One 2013:**8**(6):e66428. https://doi.org/10.1371/journal.pone.0066428

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature 2009:**461**(7263):489–494. https://doi.org/10.1038/nature08365

Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J, Dreher K, Payne T, Marshall D, et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. Nat Commun. 2020:**11**(1):4572. https://doi.org/10.1038/s41467-020-18404-w

Schulthess AW, Kale SM, Liu F, Zhao Y, Philipp N, Rembe M, Jiang Y, Beukert U, Serfling A, Himmelbach A, et al. Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. Nat Genet. 2022:**54**(10):1544–1552. https://doi.org/10.1038/s41588-022-01189-7

Shimizu KK, Copetti D, Okada M, Wicker T, Tameshige T, Hatakeyama M, Shimizu-Inatsugi R, Aquino C, Nishimura K, Kobayashi F, et al. De novo genome assembly of the Japanese wheat cultivar Norin 61 highlights functional variation in flowering time and *Fusarium*-resistance genes in East Asian genotypes. Plant Cell Physiol. 2020:**62**(1):8–27. https://doi.org/10.1093/pcp/pcaa152

Song X, Wang D, Ma L, Chen Z, Li P, Cui X, Liu C, Cao S, Chu C, Tao Y, et al. Rice RNA-dependent RNA polymerase 6 acts in small RNA biogenesis and spikelet development. Plant J. 2012:**71**(3):378–389. https://doi.org/10.1111/j.1365-313X.2012.05001.x

Sun C, Dong Z, Zhao L, Ren Y, Zhang N, Chen F. The wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnol J. 2020:**18**(6):1354–1360. https://doi.org/10.1111/pbi.13361

Sun L, Yang W, Li Y, Shan Q, Ye X, Wang D, Yu K, Lu W, Xin P, Pei Z, et al. A wheat dominant dwarfing line with *Rht12*, which reduces stem cell length and affects GA synthesis, is a 5AL terminal deletion line. Plant J. 2018:**97**(5):887–900. https://doi.org/10.1111/tpj.14168

Tian Q, Shen L, Luan J, Zhou Z, Guo D, Shen Y, Jing W, Zhang B, Zhang Q, Zhang W. Rice shaker potassium channel *OsAKT2* positively regulates salt tolerance and grain yield by mediating K^+ redistribution. Plant Cell Environ. 2021:**44**(9):2951–2965. https://doi.org/10.1111/pce.14101

Turner A, Beales J, Faure S, Dunford RP, Laurie DA. The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. Science. 2005:**310**(5750):1031–1034. https://doi.org/10.1126/science.1117619

Wang B, Lin Z, Li X, Zhao Y, Zhao B, Wu G, Ma X, Wang H, Xie Y, Li Q, et al. Genome-wide selection and genetic improvement during modern maize breeding. Nat Genet. 2020:**52**(6):565–571. https://doi.org/10.1038/s41588-020-0616-3

Wang Z, Wang W, Xie X, Wang Y, Yang Z, Peng H, Xin M, Yao Y, Hu Z, Liu J, et al. Dispersed emergence and protracted domestication of polyploid wheat uncovered by mosaic ancestral haploblock inference. Nat Commun. 2022:**13**(1):3891. https://doi.org/10.1038/s41467-022-31581-0

Wang S, Wu K, Qian Q, Liu Q, Li Q, Pan Y, Ye Y, Liu X, Wang J, Zhang J, et al. Non-canonical regulation of SPL transcription factors by a human OTUB1-like deubiquitinase defines a new plant type rice associated with higher grain yield. Cell Res. 2017:**27**(9):1142–1156. https://doi.org/10.1038/cr.2017.98

Wang D, Yu K, Jin D, Sun L, Chu J, Wu W, Xin P, Gregova E, Li X, Sun J, et al. Natural variations in the promoter of *Awn Length Inhibitor 1* (*ALI-1*) is associated with awn elongation and grain length in common wheat. Plant J. 2019:**101**(5):1075–1090. https://doi.org/10.1111/tpj.14575

Xie W, Wang G, Yuan M, Yao W, Lyu K, Zhao H, Yang M, Li P, Zhang X, Yuan J, et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. Proc Natl Acad Sci U S A. 2015:**112**(39):E5411–E5419. https://doi.org/10.1073/pnas.1515919112

Xiong H, Zhou C, Fu M, Guo H, Xie Y, Zhao L, Gu J, Zhao S, Ding Y, Li Y, et al. Cloning and functional characterization of *Rht8*, a "Green Revolution" replacement gene in wheat. Mol Plant. 2022:**15**(3): 373–376. https://doi.org/10.1016/j.molp.2022.01.014

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proc Natl Acad Sci U S A. 2006:**103**(51):19581–19586. https://doi.org/10.1073/pnas.0607142103

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 2004:**303**(5664):1640–1644. https://doi.org/10.1126/science.1094305

Yan X, Zhao L, Ren Y, Dong Z, Cui D, Chen F. Genome-wide association study revealed that the *TaGW8* gene was associated with kernel size in Chinese bread wheat. Sci Rep. 2019:**9**(1):2702. https://doi.org/10.1038/s41598-019-38570-2

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011:**88**(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zeven AC. Landraces: a review of definitions and classifications. Euphytica 1998:**104**(2):127–139. https://doi.org/10.1023/A:1018683119237

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019:**35**(10):1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Zhang L, Liu P, Wu J, Qiao L, Zhao G, Jia J, Gao L, Wang J. Identification of a novel ERF gene, *TaERF8*, associated with plant height and yield in wheat. BMC Plant Biol. 2020:**20**(1):263. https://doi.org/10.1186/s12870-020-02473-6

Zhang L, Zhang H, Qiao L, Miao L, Yan D, Liu P, Zhao G, Jia J, Gao L. Wheat MADS-box gene *TaSEP3-D1* negatively regulates heading date. Crop J. 2021:**9**(5):1115–1123. https://doi.org/10.1016/j.cj.2020.12.007

Zhao X, Guo Y, Kang L, Yin C, Bi A, Xu D, Zhang Z, Zhang J, Yang X, Xu J, et al. Population genomics unravels the Holocene history of bread wheat and its relatives. Nat Plants. 2023:**9**(3):403–419. https://doi.org/10.1038/s41477-023-01367-3

Zhao DS, Li QF, Zhang CQ, Zhang C, Yang QQ, Pan LX, Ren XY, Lu J, Gu MH, Liu QQ. *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. Nat Commun. 2018:**9**(1): 1240. https://doi.org/10.1038/s41467-018-03616-y

Zhao XL, Shi ZY, Peng LT, Shen GZ, Zhang JL. An atypical HLH protein OsLF in rice regulates flowering time and interacts with OsPIL13 and OsPIL15. N Biotechnol. 2011:**28**(6):788–797. https://doi.org/10.1016/j.nbt.2011.04.006

Zhou Y, Chen Z, Cheng M, Chen J, Zhu T, Wang R, Liu Y, Qi P, Chen G, Jiang Q, et al. Uncovering the dispersion history, adaptive evolution and selection of wheat in China. Plant Biotechnol J. 2018:**16**(1): 280–291. https://doi.org/10.1111/pbi.12770

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015:**33**(4):408–414. https://doi.org/10.1038/nbt.3096

Zhou Y, Zhao X, Li Y, Xu J, Bi A, Kang L, Xu D, Chen H, Wang Y, Wang Y-g, et al. *Triticum* population sequencing provides insights into wheat adaptation. Nat Genet. 2020:**52**(12):1412–1422. https://doi.org/10.1038/s41588-020-00722-w