Plant Communications
Resource article

CellPress
Partner Journal

Fast integration and accumulation of beneficial breeding alleles through an AB–NAMIC strategy in wheat

Chengzhi Jiao[1,2,5], Chenyang Hao[1,5], Tian Li[1,5], Abhishek Bohra[3], Lanfen Wang[1], Jian Hou[1], Hongxia Liu[1], Hong Liu[1], Jing Zhao[1], Yamei Wang[1], Yunchuan Liu[1], Zhiwei Wang[1], Xin Jing[4], Xiue Wang[2], Rajeev K. Varshney[3,*], Junjie Fu[1,*] and Xueyong Zhang[1,2,*]

[1]The National Key Facility for Crop Gene Resources and Genetic Improvement/Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2]State Key Laboratory of Crop Genetics and Germplasm Enhancement, Cytogenetics Institute, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

[3]Centre for Crop & Food Innovation, WA State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Perth, WA 6150, Australia

[4]Smartgenomics Technology Institute, Tianjin 301700, China

[5]These authors contributed equally to this article.

***Correspondence: Rajeev K. Varshney (rajeev.varshney@murdoch.edu.au), Junjie Fu (fujunjie@caas.cn), Xueyong Zhang (zhangxueyong@caas.cn)**

https://doi.org/10.1016/j.xplc.2023.100549

ABSTRACT

Wheat (*Triticum aestivum*) is among the most important staple crops for safeguarding the food security of the growing world population. To bridge the gap between genebank diversity and breeding programs, we developed an advanced backcross–nested association mapping plus inter-crossed population (AB–NAMIC) by crossing three popular wheat cultivars as recurrent founders to 20 germplasm lines from a mini core collection. Selective backcrossing combined with selection against undesirable traits and extensive crossing within and between sub-populations created new opportunities to detect unknown genes and increase the frequency of beneficial alleles in the AB–NAMIC population. We performed phenotyping of 590 AB–NAMIC lines and a natural panel of 476 cultivars for six consecutive growing seasons and genotyped these 1066 lines with a 660K SNP array. Genome-wide association studies of both panels for plant development and yield traits demonstrated improved power to detect rare alleles and loci with medium genetic effects in AB–NAMIC. Notably, genome-wide association studies in AB–NAMIC detected the candidate gene *TaSWEET6-7B* (TraesCS7B03G1216700), which has high homology to the rice *SWEET6b* gene and exerts strong effects on adaptation and yield traits. The commercial release of two derived AB–NAMIC lines attests to its direct applicability in wheat improvement. Valuable information on genome-wide association study mapping, candidate genes, and their haplotypes for breeding traits are available through WheatGAB. Our research provides an excellent framework for fast-tracking exploration and accumulation of beneficial alleles stored in genebanks.

Key words: wheat, genebank, AB–NAMIC, genome-wide association studies, GWAS, beneficial alleles, genomics-assisted breeding, WheatGAB

Jiao C., Hao C., Li T., Bohra A., Wang L., Hou J., Liu H., Liu H., Zhao J., Wang Y., Liu Y., Wang Z., Jing X., Wang X., Varshney R.K., Fu J., and Zhang X. (2023). Fast integration and accumulation of beneficial breeding alleles through an AB–NAMIC strategy in wheat. Plant Comm. **4**, 100549.

INTRODUCTION

Common wheat (*Triticum aestivum*) is the most important staple crop for safeguarding the food security of the growing human population worldwide. Wheat domestication and improvement have led to fixation of major-effect loci that control plant development and yield traits; these include semi-dwarf genes *Rht1* and *Rht2* (Zhang et al., 2006), grain size genes *GW2* and *GW6* (Su et al., 2011), and starch synthesis genes *SUS1*, *SUS2*, and *ADP1* (Hou et al., 2014, 2017).

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

Accelerating future improvements will rely on the discovery and accumulation of middle- and minor-effect genes. Systematic exploitation of the genetic and breeding potential of genebank collections is a route we must take to face the food security challenge.

How to efficiently accumulate beneficial alleles is a crucial technical question in breeding. The strategic use of diverse germplasm in pre-breeding and cultivar development programs is vital for matching the targeted rate of yield improvement (Varshney et al., 2021c). Pre-breeding in the genomics era relies on developing new approaches that accelerate exploration of novel diversity and facilitate rapid and efficient incorporation of allelic variation in crop improvement programs (Varshney et al., 2021a; Bohra et al., 2022). Genomic breeding strategies that optimize crop genomes through accumulation of beneficial alleles and removal of deleterious alleles will be indispensable for designing future crops and genomics-assisted breeding (Mascher et al., 2019; Varshney et al., 2021b).

Crop scientists have developed and implemented novel experimental designs based on multi-parent populations to incorporate vast amount of genetic diversity (Arrones et al., 2020; Scott et al., 2020). One such design, advanced backcross–nested association mapping (AB–NAM) has proven to be an efficient strategy for discovering novel variation in exotic germplasms (Nice et al., 2016; Wingen et al., 2017). Recently, it was also recognized that introgression of wild emmer and extensive selection in modern breeding have led to strong haplotype blocks (hap-blocks) across the centromeres, which heavily affect agronomic traits with rare recombination (Cheng et al., 2019; Hao et al., 2020). Integration breeding of favorable hap-blocks in different breeding systems should be a feasible strategy for breeding new cultivars.

To promote more recombination and accumulation of beneficial alleles in the population, we extended the AB–NAM strategy to more recurrent founders (RFs) with extensive crossing within and between sub-populations while imposing selection against undesirable traits. We crossed three elite cultivars with 20 genotypes from a mini core collection (Hao et al., 2008) and backcrossed twice to replace weak deleterious alleles through selection against disadvantageous breeding traits. This invaluable genetic resource is referred to as AB–NAM plus inter-crossed breeding (AB–NAMIC), and it makes novel haplotypes and genes available for genetic research and breeding. Distribution of best linear unbiased predictions (BLUPs) of phenotypic values in AB–NAMIC over 6 years clearly showed the way forward toward breeding requirements compared with a natural population panel (NP) and the three RFs. Linkage disequilibrium (LD) and nucleotide diversity analysis indicated that AB–NAMIC harbored enough genetic diversity for effective dissection of agronomic traits via genome-wide association studies (GWASs), together with NP and available re-sequencing data in wheat. We also examined global wheat cultivars to track the *TaSWEET6-7B* footprints detected by GWAS in AB–NAMIC. This also illustrated the feasibility of AB–NAMIC for gene cloning. Through AB–NAMIC, we can fully integrate gene mapping, gene cloning, and genome-associated breeding in one population.

RESULTS

Construction and phenotyping of the AB–NAMIC population in wheat

We crossed three elite wheat cultivars widely planted in wheat production, i.e., Zhoumai 18 (high yield), Zhengmai 366 (good baking quality), and Handan 6172 (excellent adaptation to temperature instability in winter and early spring) (hereafter abbreviated ZM18, ZM366, and HD6172), as RFs to 20 accessions from the mini core collection (see methods; Supplemental Table 1). The resulting F_1 hybrids were backcrossed twice with the three recurrent parents and then self-crossed for at least six generations. Individuals with the desired traits were positively selected during population development to accumulate beneficial alleles and purge deleterious alleles. We made selections to sort the population that matched the breeding objectives, such as semi-dwarfness, early flowering and maturity, higher 1000-grain weight, and increased grain number. To facilitate genome optimization through enhanced recombination opportunities, lines derived from the same or different recurrent parents were further inter-crossed and then self-crossed. The derived lines attained phenotypic stability after 10 generations from the starting crosses. A total of 1200 lines were established and retained in the original AB–NAMIC population (Figure 1A).

According to pedigree information and preliminary field performance, we chose 590 lines from the AB–NAMIC population to make the multi-environment trials more manageable and reliable (Supplemental Table 2). The AB–NAMIC panel alongside the NP panel (a collection of 476 diverse accessions; Supplemental Table 3) was continuously evaluated at the Chinese Academy of Agricultural Sciences Xinxiang Experimental Station, Henan (113.5°E, 35.2°N). A set of developmental and yield traits were measured annually for 6 years from 2014 to 2020. An end-use quality trait, the SDS-sedimentation value, was measured for 2 years. Disease resistance against powdery mildew and leaf rust was also recorded each year on a zero to five scale in natural epidemic environments. For each trait, BLUPs of the lines across multiple years were obtained and used for subsequent investigations.

Compared with the NP, the AB–NAMIC showed a significant shift toward breeding requirements, including reduced plant height and days to heading and flowering (Figure 1B). The AB–NAMIC had much lower effective tiller numbers than the NP. Grain-size-related traits (1000-grain weight, grain width, and grain thickness) and grain number shifted to higher levels in the AB–NAMIC (Figure 1B), whereas spike length and spikelet number per spike shifted somewhat lower. Phenotypic variation in the AB–NAMIC was significantly lower than that in the NP. Broad-sense heritability was also estimated across years. For each trait, the AB–NAMIC population had a higher heritability than the NP population.

The AB–NAMIC population is enriched in beneficial traits

We obtained lines with higher BLUP values for effective tiller number, grain number per spike, and 1000-grain weight in the three sub-AB–NAMIC populations compared with the three RFs. Notwithstanding the strong effects of the RFs on the derived

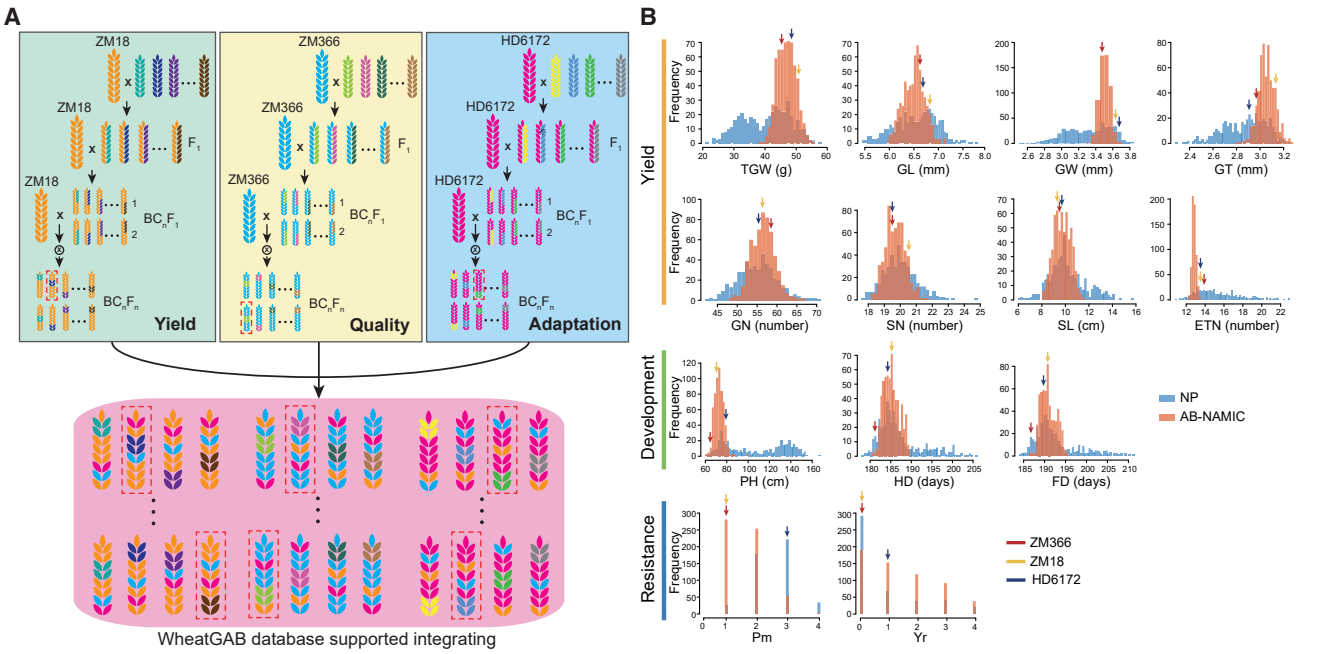

Figure 1. Strategy for AB-NAMIC construction and its advantages in agronomic traits (BLUP values).

(A) Red dotted boxes represent commercial varieties selected from the AB–NAMIC. **(B)** Blue: trait variation in NP panel; brown: trait variation in AB–NAMIC panel. TGW, 1000-grain weight (g); GL, grain length (mm); GW, grain width (mm); GT, grain thickness (mm); GN, grain number per spike; SN, spikelet number per spike; SL, spike length (cm); ETN, effective tiller number; PH, plant height (cm); HD, heading date (days); FD, flowering date (days); Pm, powdery mildew; Lr, leaf rust; Yr, yellow rust.

lines, there were lines with much higher SDS values than ZM366. ZM366 has served as a standard cultivar for checking bread-baking quality for more than 15 years by the national professional society in China. Therefore, we can infer that the mini core collection contains valuable genes (alleles) for improvement of either yield or baking quality, some of which were incorporated into our AB–NAMIC population (Supplemental Figure 1).

Differences in disease resistance among the three RFs during the six growing seasons indicated changes in pathogen races from year to year. Appearance of lines with better resistance to powdery mildew and leaf rust in the three sub-groups suggests incorporation of new resistance genes that are absent in the RFs. For instance, presence of resistant lines in the HD6172-derived sub-population strongly supported this conclusion, given that HD6172 is very sensitive to powdery mildew or leaf rust (Supplemental Figure 1).

Genomic characterization of AB–NAMIC and NP

The AB–NAMIC and NP panels were genotyped using the 660K SNP array (Affymetrix Axiom Wheat660). Of the 517 000 effective SNP markers on the 660K array, we detected 346 231 and 350 497 SNPs in the AB–NAMIC and NP, respectively (Supplemental Table 4). Most of these SNPs (345 847 SNPs) were shared by the two panels. Principal-component analysis indicated that the AB–NAMIC was separated from the NP panel, with partial overlaps (Figure 2A), whereas the separation of the three sub AB–NAM populations was less distinct because the artificial inter-crossing of excellent lines in the three AB–NAMs during population construction (Figure 2B) was consistent with the cross history among the three sub-groups (Figure 1A). The AB–NAMIC showed a decrease in diversity and an increase in genetic identity (expressed as identity by state) compared with the NP panel (Figure 2C–2E). However, an increase in LD decay from an average of 1.4 Mb in NP to 2 Mb in AB–NAMIC (distance at which the squared correlation coefficient [r^2] dropped to half of its maximum value) suggests that the decrease in recombination in the AB–NAMIC population is acceptable for subsequent genetic dissections (Figure 2F).

More substantial GWAS signals were detected in AB–NAMIC than in the NP

We carried out GWASs for traits that contribute to plant development, yield, and disease resistance in the AB–NAMIC and NP panels. In total, 2932 and 1096 peak SNPs associated with GWAS quantitative trait loci (QTLs) were detected in the AB–NAMIC and NP panels, respectively, using BLUP data (Figure 3 and Supplemental Figures 2–6; Supplemental Tables 5–7). In the NP panel, five high-confidence GWAS peaks were detected for development-related traits on chromosomes 2D, 4B, 4D, 6D, and 7D (with $-\log_{10}(p) \geq 6$) and were within LD regions of *PPD1*, *Rht-B1*, *Rht-D1*, and *VRN-D3* (*FT* homolog) (Figure 3B). For yield-related traits, *GW6*, *GN1-B1* (chr2B), *Tg1-D1* (chr2D), and several other previously known genes related to starch synthesis were found within GWAS peaks for grain number per spike (GN), grain weight (GW), and 1000-grain weight (TGW) in the NP panel (Figure 3D).

In the AB–NAMIC panel, backcrossing and selection dramatically increased the frequencies of favorable alleles of major genes

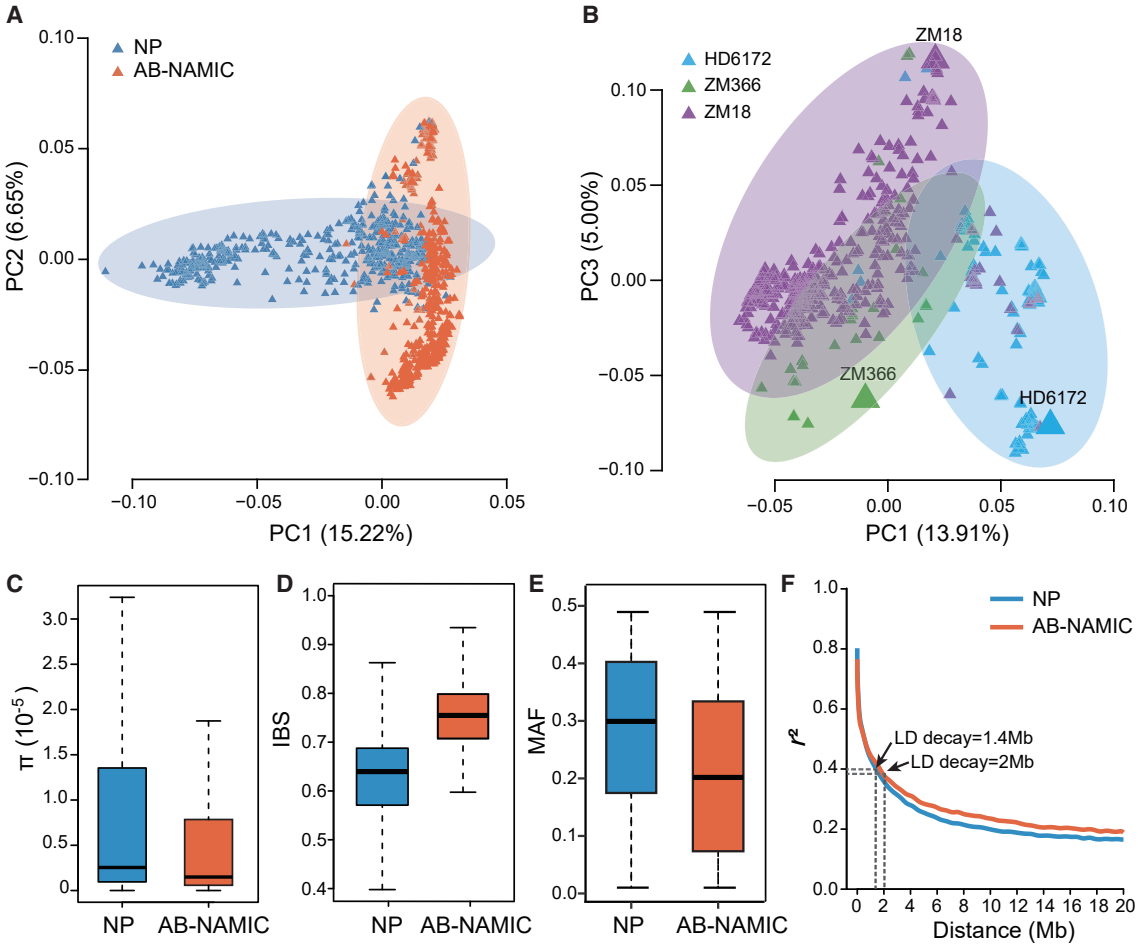

Figure 2. Phylogenetic relationships, LD values, and genetic diversity of the AB–NAMIC and NP panels.

(A) Principal-component analysis plot of all 1066 wheat accessions. Blue indicates natural panels, and orange indicates AB–NAMIC panels.

(B) Principal-component analysis plot of AB–NAMIC panels without NP panel entries.

(C–F) Genetic diversity, genetic identity (IBS), minor allele frequency, and LD decay of AB–NAMIC and NP panels. The small difference in LD decay distances between AB–NAMIC and NP indicates good recombination in the AB–NAMIC panel.

such as *PPD1*, *Rht1*, *VRN1*, and *TGW6*, leading to their fixation (Figure 3E and 3F). No or much less segregation of these major effects renders the genetic effects of other genes controlling the same traits detectable by GWASs, including genes such as *BMY-4A*, *BMY-4B*, *SSIIb*, and *TaSWEET6-7B* (Figure 3A and 3C). The increased frequency of rare alleles also facilitates detection of more association signals in the AB–NAMIC compared with the NP (Figure 3G). As expected, phenotypic variation explained (PVE) was significantly increased for developmental traits, disease resistance, and yield-related traits in AB–NAMIC compared with NP, with the exception of effective tiller number (ETN) and grain thickness (GT) (Figure 3H; Supplemental Table 8). Therefore, GWAS in AB–NAMIC offers an excellent strategy and platform for discovering novel variations by raising the frequency of rare alleles and the fixation of major genes. For wider application of functional genomic information in wheat improvement, we constructed a database (WheatGAB; www.wheatgab.com) by integrating associated SNP variations from the AB–NAMIC panel with those from the NP panel and our previous re-sequencing dataset of 145 wheat accessions (Supplemental Figure 7).

To verify the power and reliability of GWAS in the AB–NAMIC panels, we selected an interval (721 864 236–724 729 233 bp) associated with TGW in phenotyping data of BLUP and multiple environments. However, no similar GWAS association was detected for this genomic region in NP (Figure 4A and 4B). One SNP (chr7B:722880243) was associated with TGW in five environments (Supplemental Table 6). *TaSWEET6-7B* (TraesCS7B03G1216700) is the nearest gene to this peak SNP (4841 bp downstream) and has high homology to rice *SWEET6b* (Supplemental Figure 8). The re-sequencing data from 145 wheat cultivars were then used to increase SNP density (Hao et al., 2020), and two new SNPs (chr7B:722872964 and chr7B:722875395) were identified in the UTR region of the candidate *TaSWEET6-7B* (Figure 4C).These two SNPs were also confirmed by PCR (see methods). Strong LD was detected surrounding this gene in both panels (Figure 4A and 4B). The *SWEET* gene family is involved in plant development, including processes such as pollen fertility, seed setting, and grain filling (Chen et al., 2012; Ma et al., 2017; Li et al., 2021). Four haplotypes were detected for the candidate gene *TraesCS7B03G1216700*, which showed significant associations with TGW, GN, and ETN in our AB–NAMIC panel

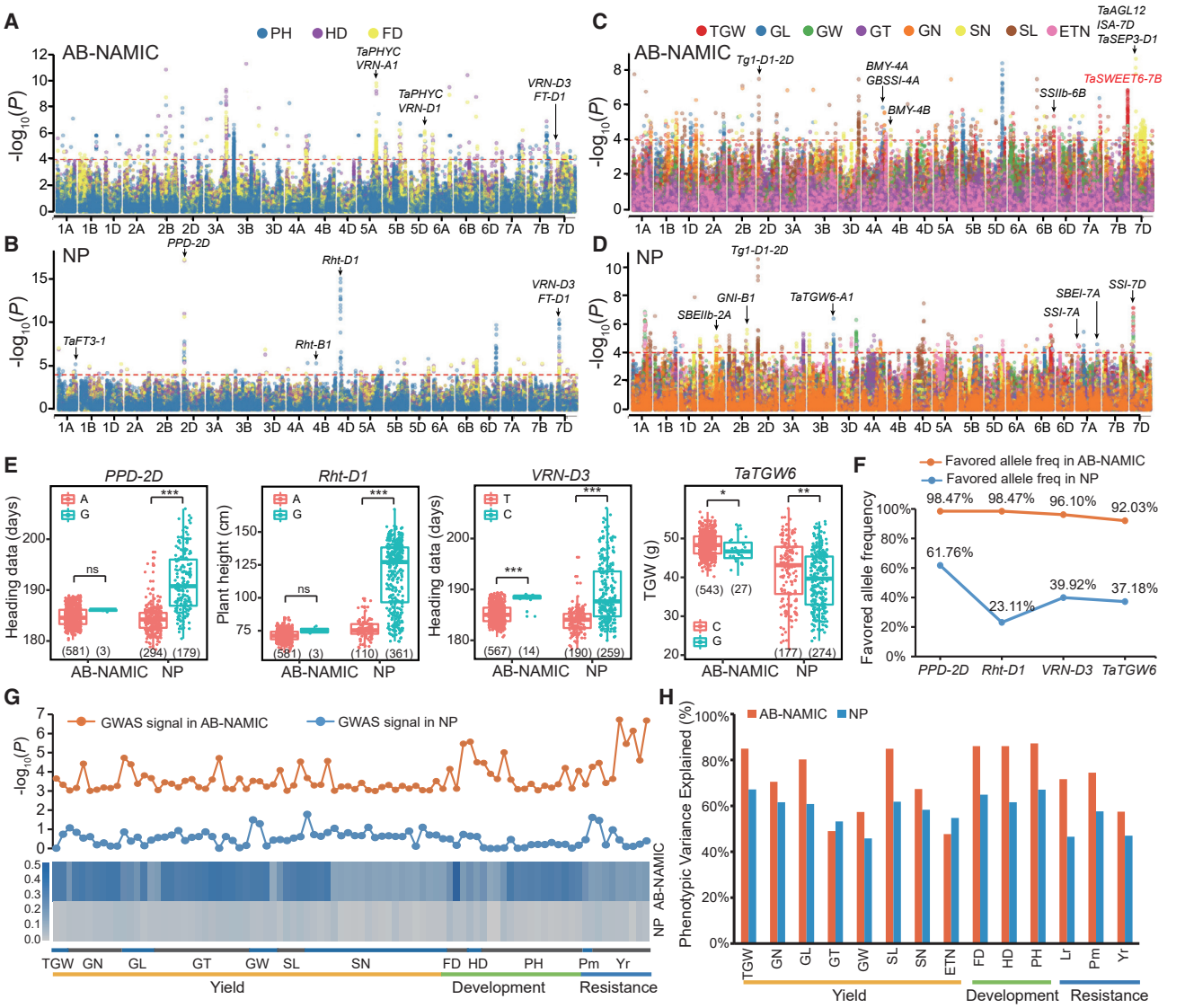

Figure 3. GWAS power for rare alleles was significantly increased in AB–NAMIC.

GWAS Manhattan plots of traits in AB–NAMIC and NP panels using BLUP data for **(A and B)** developmental traits and **(C and D)** yield traits.

(E) Estimation of the effects of well-known major genes in the AB–NAMIC and NP panels.

(F) Frequency of four major genes with strong pleiotropic effects in AB–NAMIC and NP panels. All of them were almost fixed in the AB–NAMIC panel (≥92%).

(G) Associated loci detected from the AB–NAMIC population, where frequency of rare alleles in NP (minor allele frequency <0.05) reached greater than 0.1 in AB–NAMIC through background and selection. The darkness of the blue color indicates the change in allele frequency. The orange and blue lines are the $-\log_{10}(p)$ values of 78 alleles in the AB–NAMIC and NP panels, respectively.

(H) Phenotypic variance explained (PVE) by genomic variations in AB–NAMIC and NP panels. PVE was much higher in the AB–NAMIC panel than in the NP panel for most yield, development, and resistance traits, except for ETN and GT.

(Figure 4C and 4D). Geographic distribution of the haplotypes in Chinese landraces (CLs) and modern Chinese cultivars (MCCs) revealed their genetic effects on yield-related traits (Figure 4E and 4F). For example, *Hap-4* dominated the three southern zones of China with sufficient rainfall and a highly humid environment, favoring landraces with fewer ETNs. But in the modern cultivars, *Hap-1* has taken the leading role in most regions because of positive selection on GN for higher yield. Appearance of *Hap-3* in post-2000 landmark cultivars indicated that TGW became the major trait among the three yield components for yield improvement in China. Global distribution of the four haplotypes in modern cultivars further supports the notion that *Hap-4* was favored in humid regions such as Europe, whereas *Hap-1* was favored in drought-prone regions and was found in CIMMYT and North American cultivars. High frequency of *Hap-4* in Australian cultivars may be related to the very limited yield potential in this country (Figure 4C–4H).

AB–NAMIC incorporates breeding-beneficial alleles and dramatically increases their frequency

To evaluate accumulation of alleles for breeding-beneficial traits, we considered novel associated variants from AB–NAMIC and

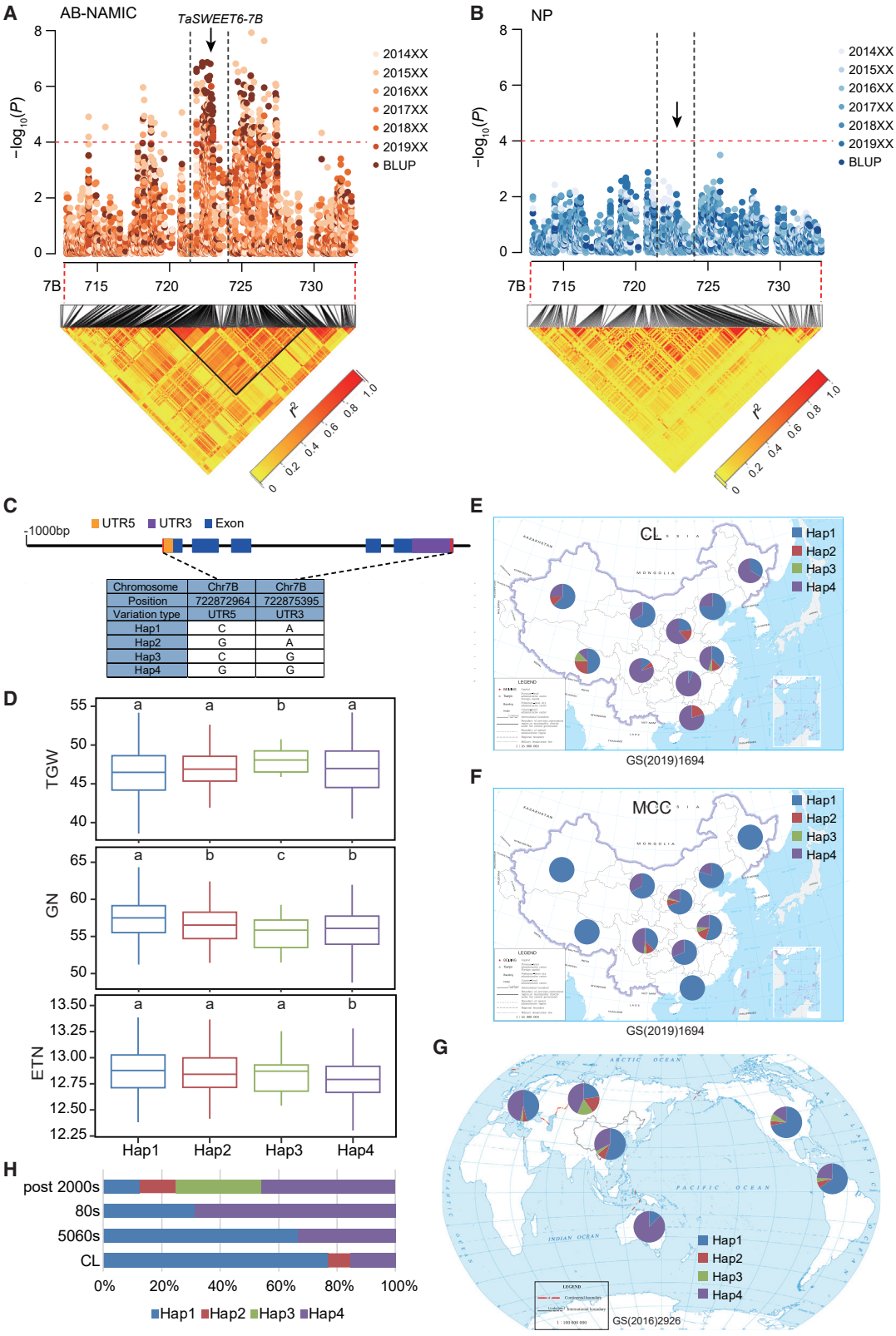

Figure 4. Geographic distribution and evolution of *TaSWEET6-7B* haplotypes.
(A and B) GWAS signals of candidate locus detected in AB–NAMIC but not in NP.
(C) The gene structure of *TaSWEET6-7B* (TraesCS7B03G1216700) and its haplotypes.

(legend continued on next page)

associated variants with relatively large effects from NP to define the ''breeding-beneficial allele'': the allele at each significant GWAS locus that had a positive effect on target traits and whose frequency increased during breeding selection (see methods). In total, 1428 beneficial alleles were identified (387 for flowering date, 453 for heading date, 171 for TGW, 77 for grain number, 90 for powdery mildew resistance, 207 for leaf rust resistance, and 43 for yellow rust resistance; Figure 5A; Supplemental Table 9). Based on identity-by-descent (IBD) analysis (Bosse et al., 2014; Wang, 2019), most beneficial alleles (1089/1428) were found to be shared by the RFs and the 20 MCC germplasms. Thirty-five beneficial alleles were identified as having been introduced from the 20 MCC germplasms (Figure 5A; Supplemental Tables 9 and 10). We subsequently used genomic prediction to examine the effective contribution of beneficial alleles for each breeding trait. One half of the AB–NAMIC panel was randomly selected as a training population 200 times to predict the genomic estimated breeding value (GEBV) of the other half based on the loci that had beneficial alleles. Compared with beneficial alleles from the NP panel, beneficial alleles from the AB–NAMIC panel significantly improved prediction accuracy for each target trait. Combining the beneficial alleles from both panels produced the highest accuracy, resulting in medium to high prediction accuracy for each target trait (0.57–0.87) (Figure 5B). These results validated the effectiveness of the detected beneficial alleles for trait improvement.

Accumulation of beneficial alleles in the AB–NAMIC was evidenced by an increase in both the number and frequency of beneficial alleles. The number of beneficial alleles correlated with their target traits in the AB–NAMIC (Figure 5C–5F and Supplemental Figure 9), implying an accumulation of beneficial alleles involved in trait improvement. We also observed a faster accumulation of beneficial alleles for target traits in the AB–NAMIC than in the national cultivars released over the last 20 years (Figure 5C–5F and Supplemental Figure 9), confirming the potential utility of the AB–NAMIC for modern wheat breeding.

The breeding value of the AB–NAMIC was also demonstrated directly by the release of two new cultivars derived from this population, Zhongmai66 and Zhongyou69, for commercial cultivation (Figure 6 and Supplemental Figure 10). Zhongyou69 is a high-yielding cultivar screened out from integration of ZM18–NAM and ZM366–NAM, whereas Zhongmai60 is descended from ZM18 (Figure 6A and 6B). Zhongmai66 was screened out from ZM18–NAM and integrated some beneficial alleles from Amenniu (Supplemental Figure 10). Based on their genotypes, analysis showed that Zhongyou69 and Zhongmai66 had accumulated more beneficial alleles than their parents for heading date (HD), GN, and powdery mildew (Pm), but not for TGW (Figure 6C and Supplemental Figure 10C). Zhongyou69 contains fewer beneficial alleles for TGW but always had a significantly higher PVE for TGW than Zhengmai366 or Zhongmai60 (Figure 6D and Supplemental Figure 10D). This implies the existence of novel major TGW alleles in Zhongyou69, probably integrated from Qiangweimai.

DISCUSSION

Rapid advances in genomics enable fast-track targeted manipulation of allelic variation to create novel diversity and facilitate its rapid and efficient integration into crop improvement programs (Tanksley and McCouch, 1997; Bevan et al., 2017; Balfourier et al., 2019; Langridge and Waugh, 2019; Milner et al., 2019; Sansaloni et al., 2020; Varshney et al., 2021b). Our AB–NAMIC strategy leverages advantages of both AB–NAM and MAGIC designs, enhancing its relevance to practical breeding and research. NAM and MAGIC designs, two typical multi-parental population designs, facilitate high-resolution mapping of genomic regions associated with complex traits through controlled crossing (Yu et al., 2008; McMullen et al., 2009; Gardner et al., 2016; Scott et al., 2020). These two classical types of synthetic population tend to avoid artificial selection in order to maintain functional loci, from major-effect to small-effect loci, while segregating into a population with balanced allele frequency. By contrast, selective backcrossing during AB–NAMIC development promotes rapid purging of deleterious alleles from the donors. Fixing indispensable alleles at major-effect loci, such as the major ''green revolution genes'' *Rht1*, *Rht2*, and *PPD1* according to local wheat production conditions, is a practical strategy in the development of AB–NAMIC lines to meet breeding needs (Figure 3). Fixing these known genes through backcrossing and medium-strength selection in turn paves the way for detecting the effects of other unknown genes (Figure 3). Analysis of the evolution of favored alleles in Chinese wheat cultivars released at different times also confirmed the breeding value of these loci (Figure 5). Inter-crosses between lines, especially lines in different sub-populations, makes it possible to incorporate desired hap-blocks derived from the three elite RFs (Hao et al., 2020). From an application standpoint, these loci and hap-blocks can easily be integrated into breeding programs through selection of elite lines in the AB–NAMIC panel, as exemplified by the release of Zhongmai66 and Zhongyou69 (Figure 6 and Supplemental Figure 10).

A distinctive multi-parental population is intended to include more genetic variation and balance allele frequencies by controlled crosses, with the aim of increasing detection power, efficiency, and predictability of marker–trait association signals (Yu et al., 2008; McMullen et al., 2009; Gardner et al., 2016; Scott et al., 2020). Our results demonstrated that it is also necessary to fix the major-effect loci in order to detect allele effects hindered at non-dominant loci in addition to increasing the frequency of rare alleles through controlled crosses. Statistically, using major-effect loci as fixed effects in the mixed model can reduce some false negatives caused by these major loci but can also introduce false positives (Segura et al., 2012). The newly obtained loci should be independently validated. Instead, our AB–NAMIC approach fixed major genes directly in

(D) Haplotype effects on three yield components estimated based on BLUP in AB–NAMIC. Statistical significance was determined using two-sided *t*-tests.
(E–G) Haplotype distribution in Chinese landraces, modern Chinese cultivars, and modern global cultivars.
(H) Haplotype frequency in landmark cultivars released at different periods in China.

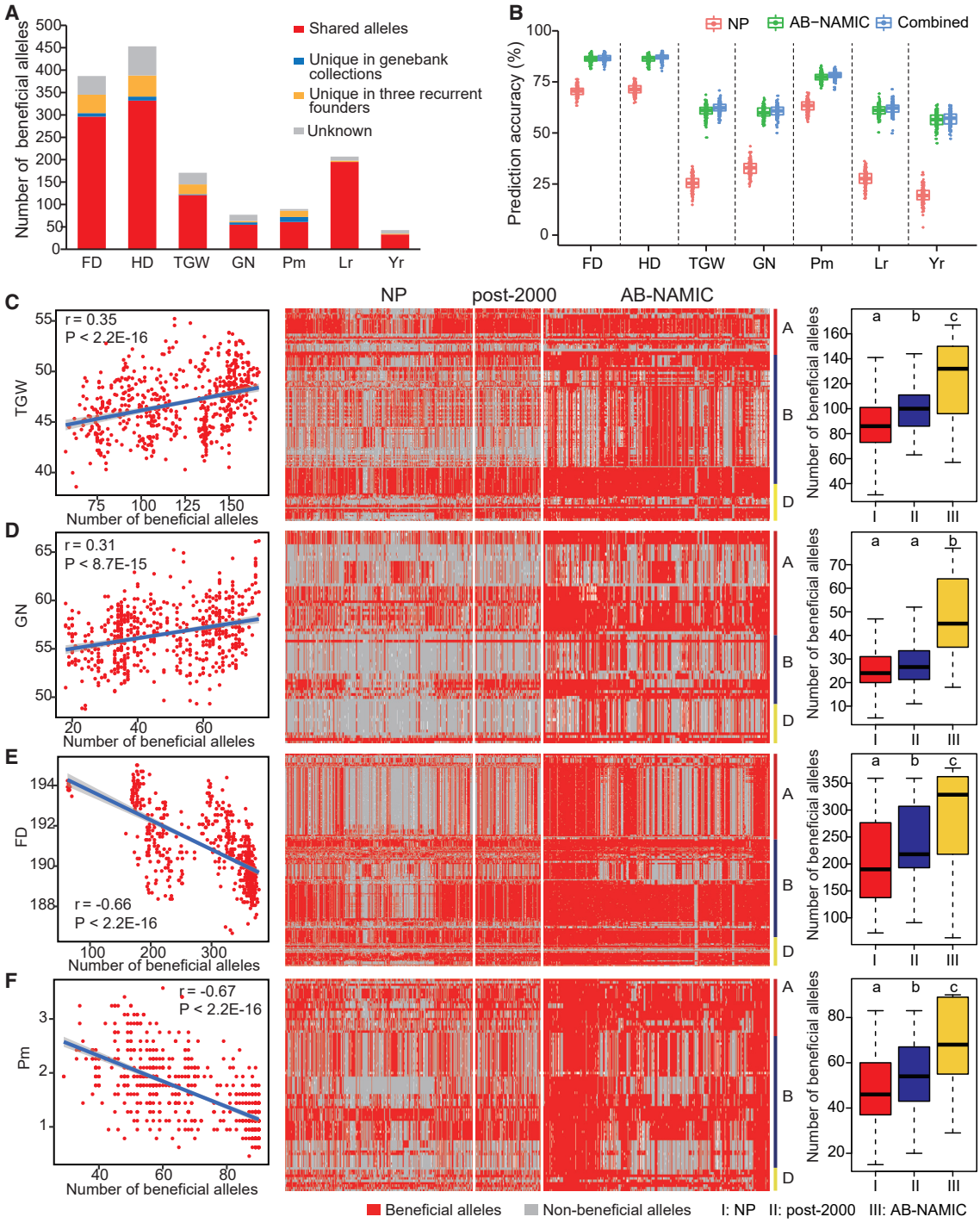

Figure 5. The AB–NAMIC lines likely accumulated a higher percentage of beneficial alleles than the NP and post-2000 lines.

(A) Sources of beneficial alleles associated with seven traits.

(B) Genomic prediction accuracies for FD, HD, TGW, GN, Pm, Lr, and Yr using beneficial alleles from GWAS peak SNPs in NP, AB–NAMIC, and the combined population. BLUP data for each trait were used to performed GS analysis with the rrBLUP package.

(C)–(F) represent TGW, GN, FD, and Pm. Left: correlation of beneficial alleles with traits. Middle: accumulation of beneficial alleles in NP, post-2000, and AB–NAMIC panels. Right: accumulation ratio of beneficial alleles was significantly higher in AB–NAMIC than in the post-2000 cultivars, indicating its utility for breeding new cultivars. The Pearson correlation coefficient (*r*) and *p* value are presented in **(C)–(F)**. Statistical significance was determined using two-sided *t*-tests.

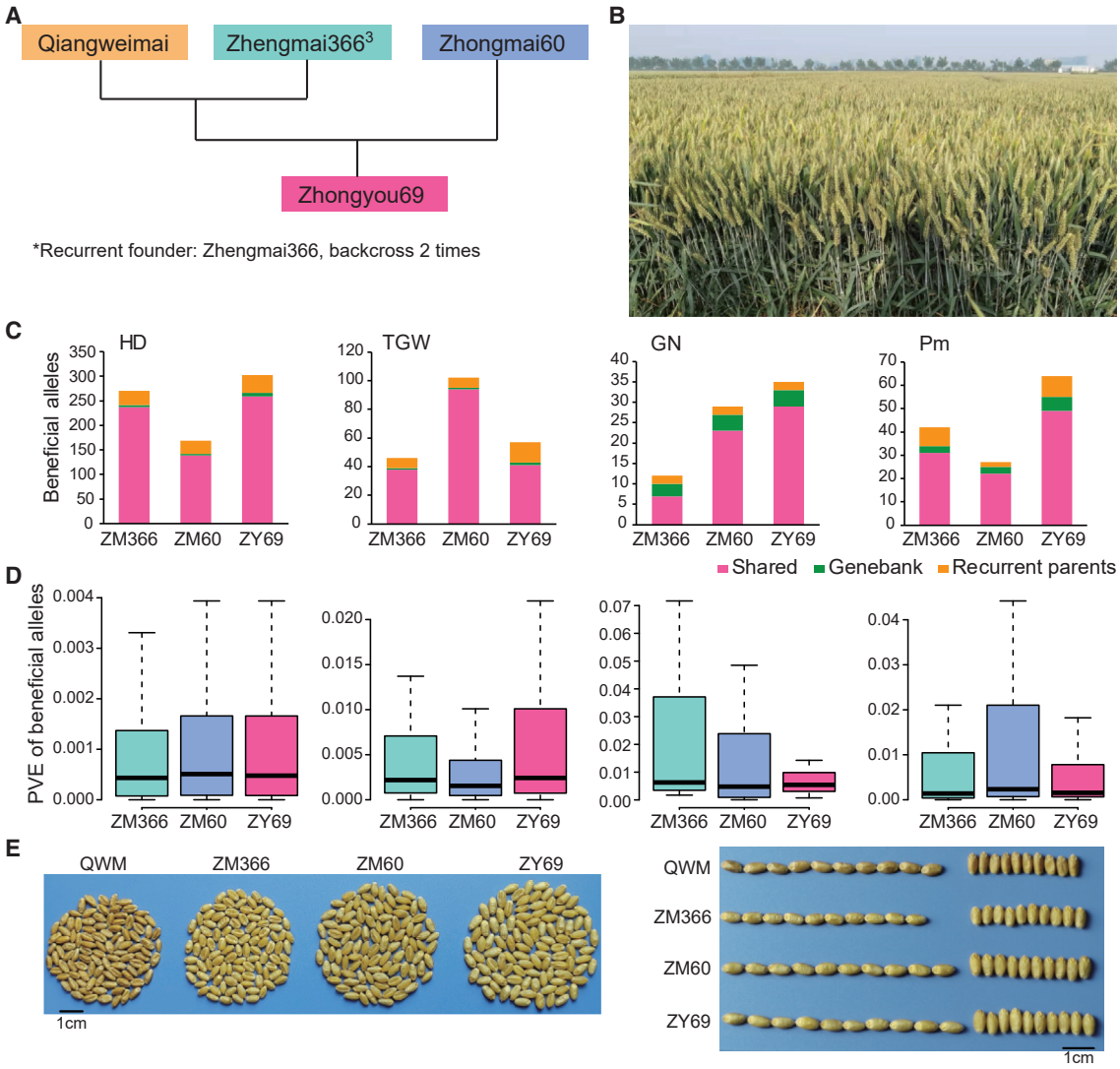

Figure 6. Accumulation of beneficial alleles in Zhongyou69, a new cultivar to be released.

(A) Family relationships of Zhongyou69.

(B) Field photograph of Zhongyou69.

(C) Statistics of beneficial alleles for HD, TGW, GN, and Pm in Zhengmai366, Zhongmai60, and Zhongyou69.

(D) PVE of beneficial alleles for HD, TGW, GN, and Pm in Zhengmai366, Zhongmai60, and Zhongyou69.

(E) Comparison of grain phenotypes among Zhengmai366, Zhongmai60, and Zhongyou69.

the population through modest selection. This approach excluded their genetic contributions to phenotypic variation in the AB–NAMIC population while largely maintaining segregation of modest- and small-effect loci. Although phenotypic variation generally decreased in the AB–NAMIC population with fixed major loci, backcross and selection during population development remained effective for increasing the detectability of non-dominant genes on behalf of increased heritability and acceptable LD decay.

It usually takes more than 10 years to clone a target gene in wheat by fine mapping in a bi-parental population, even for genes with dominant effects on fertility or resistance (Xia et al., 2017; Lu et al., 2020). By high-density genotyping of the AB–NAMIC panel, we delineated candidate gene haplotypes and demonstrated their contributions to historical breeding decisions (Figure 4). Notably, *Hap-4* at *TaSWEET6-7B* was validated as a novel haplotype for higher yield that increased TGW, and it can be directly deployed into a breeding program through the AB–NAMIC lines.

The WheatGAB database established in this study can facilitate functional investigation by integrating raw GWAS signals, their functional annotations, and external information such as previously detected QTLs. We anticipate that further whole-genome re-sequencing of AB–NAMIC lines will accelerate mapping of complex agronomic traits that are crucial for wheat improvement (Yano et al., 2016; Milner et al., 2019) (Figure 3 and Supplemental Figures 2–6).

In summary, the AB–NAMIC has fully merged functional gene discovery and selective breeding into one population in wheat. Our strategy overcomes the problem of unintended phenotypic

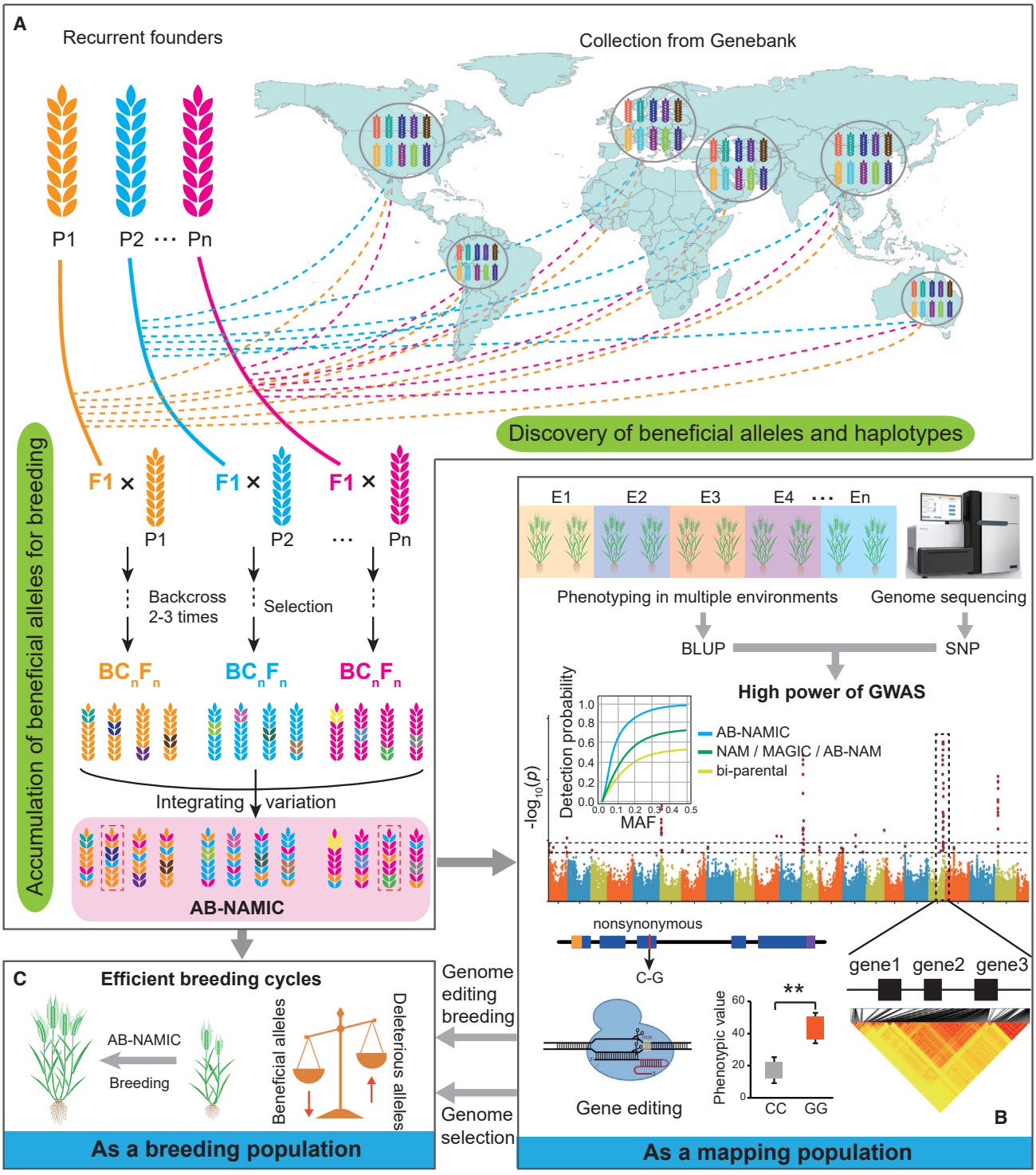

Figure 7. Scheme for unlocking genes hidden in genebank collections by AB–NAMIC and integrating them into a breeding program.
(A) Discovery and accumulation of beneficial alleles from genebank collections by the AB–NAMIC strategy.
(B) As a mapping population, AB–NAMIC offers an excellent platform for discovering novel variations by increasing the frequency of rare alleles and fixing major genes.
(C) As a breeding population, AB–NAMIC incorporates breeding-beneficial alleles for creation of excellent varieties.

outcomes that are often encountered when QTL discovery and transfers are performed in different genetic backgrounds. Release of new cultivars derived from this population directly demonstrated the feasibility of AB–NAMIC in breeding. AB–NAMIC is an excellent strategy for integrating germplasm diversity, genomics, and breeding to usher in the breeding 4.0 era. The AB–NAMIC strategy is suitable not only for wheat but also for all self-crossing crops (Figure 7).

METHODS

Natural GWAS panel

A natural population (NP) of 476 common wheat accessions, including 158 CLs, 298 MCCs, and 20 introduced modern cultivars, was used in the current study. Except for the most recent modern cultivars released after the 2000s in China, more than half of the materials were from a Chinese wheat mini core collection that contains 1% of Chinese wheat initial accessions but represents more than 70% of the total genetic diversity (Hao et al., 2008). The MCCs included cultivars released in the 1950s (18 collections), 1960s (19), 1970s (27), 1980s (41), 1990s (46), and post-2000 (147). Detailed information on the 476 accessions is given in Supplemental Table 2.

Construction of the AB–NAMIC

An AB–NAMIC population was successfully constructed in the current study. Twenty accessions, mainly from Chinese core collections, were crossed as donors with three elite MCCs, i.e., ZM18, ZM366, and HD6172, as RFs in 2004. The derived F_1 plants were backcrossed twice with the three recurrent parents and then self-crossed for at least six generations. Individual plants with superior breeding traits, such as semi-dwarfism, earlier flowering and maturity, higher TGW, higher GN, etc., were retained to accumulate advantageous alleles and remove deleterious alleles as much as possible through artificial selection of desired phenotypes. In addition, superior lines generated from different donors but with the same RF, as well as lines from the three different RFs, were crossed with each other to produce more recombination events and increase the genetic diversity of the population. Overall, 1200 derived lines of the AB–NAMIC population reached morphological stability by 2013 and were phenotyped over subsequent years. Then, 590 lines were selected for genotyping with the 660K SNP array. The strategy for construction of the AB–NAMIC population is shown in Figure 1. Detailed information on these lines from the AB–NAMIC population is provided in Supplemental Table 3.

Phenotypic measurement and statistical analysis

The 476 NP accessions and the 590 AB–NAMIC lines were all planted at the Chinese Academy of Agricultural Sciences Xinxiang Experimental Station, Henan (113.5°E, 35.2°N) in six growing seasons from 2014 to 2020. On the basis of their planting years, they were named 2014XX, 2015XX, 2016XX, 2017XX, 2018XX, and 2019XX, respectively. Each NP accession was planted in a 4-m four-row plot with 25 cm between rows and 40 seeds per row. Field management followed standard practices. For NP and AB–NAMIC populations, 10 plants from the middle of each plot were measured to phenotype agronomic traits, including HD (day); flowering date (FD; day); spike length (SL, cm); spikelet number per spike; plant height (cm); GN; ETN; TGW (g); grain length (mm); grain width (mm); and GT (mm), as well as resistance to some diseases, such as Pm, leaf rust (Lr), and yellow rust (Yr), identified in the natural environments. The disease severity was scored from 0–4, indicating responses from immunity to serious infection. Micro SDS-sedimentation value (mL) using 2 g wheat flour was evaluated as an important quality trait for each line of the AB–NAMIC population.

Trait BLUP value estimation

The BLUP values for each line in the NP and AB–NAMIC populations were estimated across multiple years using a mixed linear model with the lme4 package in R (v.3.2.2). The formula was as follows: $Y = \mu + Line + Loc + (Line \times Loc) + Rep\ (Loc) + \varepsilon$. Y represents the phenotype; *Line* and *Loc* denote the respective random effects of the genotype and the environment; *Line* × *Loc* is the interaction effect between genotype and environment; *Rep* (*Loc*) is the random effect of replication nested in the environment; and ε represents the random residual error effect. The BLUP values were used to analyze the phenotypic distribution in NP and AB–NAMIC populations. Significant differences in phenotypic data were assessed using *t*-tests at significance levels of 0.05 or 0.01.

Genotyping, filtering, and annotation of the raw data

Genomic DNA of each material was extracted from fresh leaves of individual 10-day-old seedlings (100 mg) using the DNAquick Plant System (Tiangen Biotech, Beijing, China; www.tiangen.com) according to the manufacturer's instructions. The DNA was diluted to a final concentration of 40 ng μl^{-1} for use in genotyping.

The Affymetrix Axiom Wheat660 array (Sun et al., 2020) composed of 630 517 SNPs covering the entire genome (http://wheat.pw.usda.gov/ggpages/topics/Wheat660_SNP_array_developed_by_CAAS.pdf) was used to perform genotyping of the two panels by CapitalBio Corporation (http://www.capitalbio.com), and data capture was performed using the Affymetrix Analysis Suite following the manufacturer's instructions with DQC >0.6 and QC call rate >70% together with manual filtering. The final available numbers of polymorphic SNPs in the NP and AB–NAMIC populations were 350 497 and 346 231, respectively, after filtering with the following parameters: minor allele frequency ≥ 0.05, with missing rate ≤ 0.2. The identified SNPs were further annotated using ANNOVAR software (v.2013-05-20) (Wang et al., 2010) and thus divided into the following groups on the basis of the wheat CS RefSeq v.2.1 (Zhu et al., 2021) genome annotation: variations occurring in intergenic regions, within 1 kb upstream (downstream) of transcription start (stop) sites; in coding sequences; and in introns.

Population structure and genetics analysis

Principal-component analysis of all 1066 accessions and 590 lines in AB–NAMIC was performed with GCTA software (Yang et al., 2011). First, we obtained the genetic relationship matrix with the parameter "-make-grm". Then, the top three principal components were estimated with the parameter "-pca3".

Nucleotide diversity was used to estimate the degree of variability within each panel (NP and AB–NAMIC) by VariScan (v.2.0.3) (Vilella et al., 2005) with a window size of 1 Mb and a step of 500 kb. Identity by state was used to explain the kinship of individuals within a group with PLINK (Purcell et al., 2007) using a window size of 1 Mb and a step of 500 kb. To estimate and compare the patterns of LD for different groups, the r^2 between pairwise SNPs was computed using PLINK (Purcell et al., 2007) with the parameters "–ld-window-r2 0 –ld-window 99 999 –ld-window-kb 20 000". The average r^2 value was calculated for pairwise markers in a 10-kb window, and values were averaged across the whole genome. The LD decay distance was measured as the physical distance at which the average pairwise r^2 dropped to half of its maximum value (Huang et al., 2010). The physical distance of LD decay was 1400 kb in the NP panel and 2000 kb in the AB–NAMIC panel.

GWAS

GWAS was performed for 14 phenotypic traits, including HD, FD, PH, ETN, SL, spikelet number per spike, GN, TGW, grain length, grain width, GT, and resistance to Pm, Lr, and Yr. The NP and AB–NAMIC panels were used to perform GWASs with the EMMAX software package (Kang et al., 2010) using both BLUP values and single-environment phenotypes. The top three principal components were used to build up the Q matrix for the population-structure correction with GCTA software (Yang et al., 2011). The genetic relationship between individuals was modeled as a random effect using the kinship matrix with EMMAX (Kang et al., 2010).

GWAS-associated peak SNPs were defined using the criterion $-\log_{10}(p) \geq 4$ for BLUP-GWAS or $-\log_{10}(p) \geq 4$ in at least two environments. The candidate genomic regions were defined as the 1400- and 2000-kb regions surrounding GWAS signal peaks from the NP and AB–NAMIC panels, respectively.

Significant SNPs detected in the same LD block were merged into a single QTL. If two SNPs flanked a QTL region, the distance between the SNPs was considered to be the QTL interval. Pairwise LD of SNPs surrounding each QTL was visualized using the R package LDheatmap (Shin et al., 2006). The annotated genes within the physical intervals of identified QTLs were extracted from the IWGSC Reference Genome v.2.1 (Zhu et al., 2021).

Peak SNPs ($-\log_{10}(p) \geq 4$) identified in NP and AB–NAMIC panels using the BLUP values of phenotypic traits were used to estimate single allelic effects by the MLM model in the GAPIT package (Zhang et al., 2010).

Determination of genome-wide GWAS threshold

To determine the threshold for GWAS analysis, seven traits (ETN, TGW, SL, HD, PH, Lr, and Pm) were selected for testing. We randomly shuffled the observed real phenotypes to break the connections between these phenotypes and their corresponding genotypes. Then, we performed GWAS analysis on the putative phenotypes 1000 times randomly using the same MLM model with the EMMAX software package. The most significant *p* value across the whole genome was recorded. The distribution of the most significant *p* values across the 1000 replicates was used to determine the threshold, which was the *p* value corresponding to a 5% chance of a type I error. Ideally, each trait should have its own threshold. We found that the thresholds were higher in the NP (average = 3.409) than the AB–NAMIC population (average = 3.239) for seven phenotypes. To improve the accuracy of GWAS analysis, we used a more stringent threshold ($p < 10^{-4}$) as the criterion for the 14 traits. Although this criterion may have caused false negatives, it was believed that the type I error was below 5% for each trait.

For estimation of PVE by SNPs in NP and AB–NAMIC, we used the genomic-relatedness-based restricted maximum likelihood (GREML) method in GCTA software (v.1.24.2) (Yang et al., 2011) to estimate the ratio of genetic variance to phenotypic variance of 14 traits using genome-wide markers.

Identification of beneficial alleles

All peak SNPs with $-\log_{10}(p) \geq 4$ for BLUP-GWAS or $-\log_{10}(p) \geq 4$ in at least 2 years associated with TGW, GN, FD, HD, Pm, Lr, and Yr in AB–NAMIC and NP panels were used to identify beneficial alleles. We first calculated the effect of each allele on the phenotype in AB–NAMIC panels and removed false-positive loci (*t*-test, $p > 0.05$). Second, we defined the beneficial breeding alleles at the associated loci based on at least one of three phenotypic characteristics, i.e., positive effect on GN or TGW, earlier heading and flowering, or resistance to epidemic disease pathogens. Third, we calculated the beneficial allele frequency in NP, post-2000, and AB–NAMIC panels; if the allele frequency was significantly higher in post-2000 or AB–NAMIC than in NP, it was regarded as a beneficial allele.

Identification of beneficial alleles from genebank collections

We next estimated the possible sources of the 1428 beneficial alleles. We detected pairwise shared haplotypes between the AB–NAMIC group and one of the two other wheat groups (RFs or genebank collections [GCs]) using IBD following an approach described previously (Bosse et al., 2014; Wang, 2019; Hao et al., 2020) with minor modifications. The approach involves two main steps: identification of pairwise IBD regions and calculation of relative IBD (rIBD) statistics. Initially, all individuals were phased with the fastPHASE function in Beagle v.4.1 (Browning and Browning, 2007). Pairwise shared haplotypes were extracted with the Beagle RefinedIBD function (Browning and Browning, 2013). To profile the frequency of shared haplotypes along individual chromosomes, each chromosome was divided into 2000-kb bins, and the numbers of recorded IBD tracts between AB–NAMIC and the two groups of accessions (RF and GC) were tallied per bin. Because each group contained a different number of samples, the number of pairwise comparisons also differed between the groups, and thus the nIBD values of each bin were normalized from zero (no IBD detected between individuals within two groups) to one (complete IBD detected between individuals within two groups). The normalized IBD between the accession and the RF group ($nIBD_{RF}$) and the normalized IBD between this accession and the GC group ($nIBD_{GC}$) were then used to calculate the rIBD ($rIBD = nIBD_{GC} - nIBD_{RF}$). The source region was defined if $\geq$100 IBD paired supported. On the basis of rIBD values, we divided the genome of AB–NAMIC into four parts: shared region between RF and GC, unique in GC, unique in RF, and unknown regions. Finally, 1428 beneficial alleles were anchored into the four regions.

Genomic prediction

The 1428 beneficial alleles and the BLUP data for seven traits were used for genomic prediction with the R package rrBLUP (Endelman, 2011). In addition, 1428 beneficial alleles of each trait were divided into three sources: 114 beneficial alleles from the NP panel, 1314 from the AB–NAMIC panel, and all 1428 from the NP and AB–NAMIC panels. Beneficial alleles from three sources were used as training SNP data, and one half of the AB–NAMIC panels were randomly selected as the training population 200 times to predict the GEBV of the other half. Prediction accuracy was based on Pearson correlations between the GEBV and the observed phenotype values.

Haplotypes and functional markers for *TaSWEET6-7B*

Haplotypes of the candidate gene *TaSWEET6-7B* (TraesCS7B03G1216700) were identified based on polymorphic sequences in our previous re-sequencing data from 145 landmark cultivars (Hao et al., 2020). Functional markers for *TaSWEET6-7B* were developed based on SNPs 79C/G (5′ UTR) and 2353A/G (3′ UTR). Genome-specific primers were designed for these two SNP markers and included SWEET7B-79C/G-5′: CATCGGC TCGCTCCACTGT, SWEET7B-79C/G-3′: GCAACAAGCACCATCCATGC, SWEET7B-2353C/G-5′: GCCGTCGCCAAGTGATTCTA, and SWEET7B-2353C/G-3′: GGTTCCACAAAAACAACCGC. PCR was performed in total volumes of 15 μl that included 80 ng genomic DNA, 7.5 μl 2× Ex Taq MasterMix (CWBIO), and 10 μM of each primer. PCR amplification was performed as follows: 95°C for 3 min; 38 cycles of 95°C for 30 s, annealing at 58°C–60°C for 30 s, and extension at 72°C (40–60 s) for 30 s; and final extension at 72°C for 10 min. Finally, digestions were performed in total volumes of 10 μl that included 5 μl PCR products, 1 μl 10× CutSmart Buffer (NEB), 0.1 μl restriction enzyme (BsmI for SNP 79C/G and EcoRV for SNP 2353A/G), and 3.9 μl ddH_2O, according to the manufacturer's instructions.

The geographic distribution of the major haplotypes (Hap1, Hap2, Hap3, and Hap4) at *TaSWEET6-7B* was analyzed for 158 CLs and 298 MCCs, as well as five major wheat production regions worldwide, including North America (490), Europe (384), Australia (51), CIMMYT (53), and the former USSR (83).

Implementation and availability of the WheatGAB database

WheatGAB is implemented using MySQL (https://www.mysql.com/) and the Flask development server (https://www.w3cschool.cn/flask/; a lightweight web server written purely in Python). Web user interfaces were developed using Django (https://www.djangoproject.com; a high-level Python web framework that encourages rapid development and clean, pragmatic design), HTML5, CSS3, Asynchronous JavaScript and XML (a set of web development techniques for creating asynchronous applications without interfering with the display and behavior of the existing page), and JQuery (a cross-platform and feature-rich JavaScript library; http://jquery.com, v.1.10.2), as well as Bootstrap (an open-source toolkit for developing web projects with HTML, CSS, and JS; https://getbootstrap.com, v.4.6.0). For dynamic genome visualization and analysis, JQuery was used to generate interactive charts.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This project was supported by the National Key Research and Development Program of China (2022YFD1201503 and 2016YFD0100302) and the National Major Agricultural Science and Technology Project (NK2022060101).

AUTHOR CONTRIBUTIONS

X.Z. together with J.F. and R.K.V. designed and coordinated the project. C.H., L.W., J.H., Hongxia Liu, Hong Liu, J.Z., Y.W., Y.L., and Z.W. collected the phenotyping data. C.J., J.F., C.H., A.B., and X.J. interpreted the data. T.L. performed gene cloning and haplotyping. X.Z., J.F., A.B., and R.K.V. wrote the manuscript.

ACKNOWLEDGMENTS

R.K.V. is grateful to the Food Futures Institute, Murdoch University, for financial support. No conflict of interest is declared.

Received: October 7, 2022
Revised: December 26, 2022
Accepted: January 11, 2023
Published: January 14, 2023

REFERENCES

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M.J., Prohens, J., and Gramazio, P. (2020). The dawn of the age of multi-parent MAGIC populations in plant breeding: novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. Biology **9**:229. https://doi.org/10.3390/biology9080229.

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

Bevan, M.W., Uauy, C., Wulff, B.B.H., Zhou, J., Krasileva, K., and Clark, M.D. (2017). Genomic innovation for crop improvement. Nature **543**:346–354. https://doi.org/10.1038/nature22011.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S.R., and Varshney, R.K. (2022). Reap the crop wild relatives for breeding future crops. Trends Biotechnol. **40**:412–431. https://doi.org/10.1016/j.tibtech.2021.08.009.

Bosse, M., Megens, H.J., Frantz, L.A.F., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R.P.M.A., et al. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nat. Commun. **5**:4392. https://doi.org/10.1038/ncomms5392.

Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics **194**:459–471. https://doi.org/10.1534/genetics.113.150029.

Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. **81**:1084–1097. https://doi.org/10.1086/521987.

Chen, L.Q., Qu, X.Q., Hou, B.H., Sosso, D., Osorio, S., Fernie, A.R., and Frommer, W.B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science **335**:207–211. https://doi.org/10.1126/science.1213351.

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

Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome **4**:250–255. https://doi.org/10.3835/plantgenome2011.08.0024.

Gardner, K.A., Wittern, L.M., and Mackay, I.J. (2016). A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. Plant Biotechnol. J. **14**:1406–1417. https://doi.org/10.1111/pbi.12504.

Hao, C., Dong, Y., Wang, L., You, G., Zhang, H., Ge, H., Jia, J., and Zhang, X. (2008). Genetic diversity and construction of core collection in Chinese wheat genetic resources. Chinese Sci. Bull. **53**:1518–1526.

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. https://doi.org/10.1016/j.molp.2020.09.001.

Hou, J., Jiang, Q., Hao, C., Wang, Y., Zhang, H., and Zhang, X. (2014). Global selection on sucrose synthase haplotypes during a century of wheat breeding. Plant Physiol. **164**:1918–1929. https://doi.org/10.1104/pp.113.232454.

Hou, J., Li, T., Wang, Y., Hao, C., Liu, H., and Zhang, X. (2017). ADP-glucose pyrophosphorylase genes, associated with kernel weight, underwent selection during wheat domestication and breeding. Plant Biotechnol. J. **15**:1533–1543. https://doi.org/10.1111/pbi.12735.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42**:961–967. https://doi.org/10.1038/ng.695.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**:348–354. https://doi.org/10.1038/ng.548.

Langridge, P., and Waugh, R. (2019). Harnessing the potential of germplasm collections. Nat. Genet. **51**:200–201. https://doi.org/10.1038/s41588-018-0340-4.

Li, P., Wang, L., Liu, H., and Yuan, M. (2022). Impaired SWEET-mediated sugar transportation impacts starch metabolism in developing rice seeds. Crop J. **10**:98–108. https://doi.org/10.1016/j.cj.2021.04.012.

Lu, P., Guo, L., Wang, Z., Li, B., Li, J., Li, Y., Qiu, D., Shi, W., Yang, L., Wang, N., et al. (2020). A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. Nat. Commun. **11**:680. https://doi.org/10.1038/s41467-020-14294-0.

Ma, L., Zhang, D., Miao, Q., Yang, J., Xuan, Y., and Hu, Y. (2017). Essential role of sugar transporter OsSWEET11 during the early stage of rice grain filling. Plant Cell Physiol. **58**:863–873. https://doi.org/10.1093/pcp/pcx040.

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J.C., and Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. Nat. Genet. **51**:1076–1081.

McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., et al. (2009). Genetic properties of the maize nested association mapping population. Science **325**:737–740. https://doi.org/10.1126/science.1174320.

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

Nice, L.M., Steffenson, B.J., Brown-Guedira, G.L., Akhunov, E.D., Liu, C., Kono, T.J.Y., Morrell, P.L., Blake, T.K., Horsley, R.D., Smith, K.P., et al. (2016). Development and genetic characterization of an advanced backcross-nested association mapping (AB-NAM) population of wild × cultivated barley. Genetics **203**:1453–1467. https://doi.org/10.1534/genetics.116.190736.

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

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

Scott, M.F., Ladejobi, O., Amer, S., Bentley, A.R., Biernaskie, J., Boden, S.A., Clark, M., Dell'Acqua, M., Dixon, L.E., Filippi, C.V., et al. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity **125**:396–416. https://doi.org/10.1038/s41437-020-0336-6.

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q., and Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. **44**:825–830. https://doi.org/10.1038/ng.2314.

Shin, J.H., Blay, S., Graham, J., and McNeney, B. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J. Stat. Softw. **16**. https://doi.org/10.18637/jss.v016.c03.

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). Theor. Appl. Genet. **122**:211–223. https://doi.org/10.1007/s00122-010-1437-z.

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. https://doi.org/10.1111/pbi.13361.

Tanksley, S.D., and McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. Science **277**:1063–1066. https://doi.org/10.1126/science.277.5329.1063.

Varshney, R.K., Bohra, A., Roorkiwal, M., Barmukh, R., Cowling, W.A., Chitikineni, A., Lam, H.M., Hickey, L.T., Croser, J.S., Bayer, P.E., et al. (2021a). Fast-forward breeding for a food-secure world. Trends Genet. **37**:1124–1136. https://doi.org/10.1016/j.tig.2021.08.002.

Varshney, R.K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M.E. (2021b). Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci. **26**:631–649. https://doi.org/10.1016/j.tplants.2021.03.010.

Varshney, R.K., Barmukh, R., Roorkiwal, M., Qi, Y., Jana, K., Tuberosa, R., Reynolds, M.P., Tardieu, F., and Siddique, K.H.M. (2021c). Breeding custom-designed crops for improved drought adaptation. Adv. Genet. **2**:e202100017. https://onlinelibrary.wiley.com/doi/full/10.1002/ggn2.202100017.

Vilella, A.J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005). VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics **21**:2791–2793. https://doi.org/10.1093/bioinformatics/bti403.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. **38**:e164. https://doi.org/10.1093/nar/gkq603.

Wang, X., Chen, L., and Ma, J. (2019). Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. Genome Biol. **20**:22.

Wingen, L.U., West, C., Leverington-Waite, M., Collier, S., Orford, S., Goram, R., Yang, C.Y., King, J., Allen, A.M., Burridge, A., et al. (2017). Wheat landrace genome diversity. Genetics **205**:1657–1676. https://doi.org/10.1534/genetics.116.194688.

Xia, C., Zhang, L., Zou, C., Gu, Y., Duan, J., Zhao, G., Wu, J., Liu, Y., Fang, X., Gao, L., et al. (2017). A TRIM insertion in the promoter of Ms2 causes male sterility in wheat. Nat. Commun. **8**:15407. https://doi.org/10.1038/ncomms15407.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88**:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat. Genet. **48**:927–934. https://doi.org/10.1038/ng.3596.

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178**:539–551. https://doi.org/10.1534/genetics.107.074245.

Zhang, X., Yang, S., Zhou, Y., He, Z., and Xia, X. (2006). Distribution of the Rht-B1b, Rht-D1b and Rht8 reduced height genes in autumn-sown Chinese wheats detected by molecular markers. Euphytica **152**:109–116. https://doi.org/10.1007/s10681-006-9184-6.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **42**:355–360. https://doi.org/10.1038/ng.546.

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J.C., Deal, K.R., De Oliveira, R., Choulet, F., Keeble-Gagnère, G., Tibbits, J., Rogers, J., et al. (2021). Optical maps refine the bread wheat Triticum aestivum cv. Chinese Spring genome assembly. Plant J. **107**:303–314. https://doi.org/10.1111/tpj.15289.