

Whole-genome resequencing of the wheat A subgenome progenitor *Triticum urartu* provides insights into its demographic history and geographic adaptation

Xin Wang^{1,3,7}, Yafei Hu^{4,7}, Weiming He^{4,7}, Kang Yu^{4,6,7}, Chi Zhang⁴, Yiwen Li¹, Wenlong Yang¹, Jiazhu Sun¹, Xin Li¹, Fengya Zheng⁴, Shengjun Zhou³, Lingrang Kong⁵, Hongqing Ling¹, Shancen Zhao^{4,6,*}, Dongcheng Liu^{2,*} and Aimin Zhang^{1,2,*}

¹State Key Laboratory of Plant Cell and Chromosome Engineering, National Center for Plant Gene Research, Institute of Genetics and Developmental Biology/Innovation Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

²State Key Laboratory of North China Crop Improvement and Regulation, College of Agronomy, Hebei Agricultural University, Baoding 071001, China

³Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

⁴BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

⁵State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Agronomy, Shandong Agricultural University, Tai'an 271018, China

⁶BGI Institute of Applied Agriculture, BGI-Agro, Shenzhen 518120, China

⁷These authors contributed equally to this article.

*Correspondence: Shancen Zhao (zhaoshancen@genomics.cn), Dongcheng Liu (liudongcheng@hebau.edu.cn), Aimin Zhang (amzhang@genetics.ac.cn)

<https://doi.org/10.1016/j.xplc.2022.100345>

ABSTRACT

Triticum urartu is the progenitor of the A subgenome in tetraploid and hexaploid wheat. Uncovering the landscape of genetic variations in *T. urartu* will help us understand the evolutionary and polyploid characteristics of wheat. Here, we investigated the population genomics of *T. urartu* by genome-wide sequencing of 59 representative accessions collected around the world. A total of 42.2 million high-quality single-nucleotide polymorphisms and 3 million insertions and deletions were obtained by mapping reads to the reference genome. The ancient *T. urartu* population experienced a significant reduction in effective population size (N_e) from $\sim 3\,000\,000$ to $\sim 140\,000$ and subsequently split into eastern Mediterranean coastal and Mesopotamian-Transcaucasian populations during the Younger Dryas period. A map of allelic drift paths displayed splits and mixtures between different geographic groups, and a strong genetic drift towards hexaploid wheat was also observed, indicating that the direct donor of the A subgenome originated from northwestern Syria. Genetic changes were revealed between the eastern Mediterranean coastal and Mesopotamian-Transcaucasian populations in genes orthologous to those regulating plant development and stress responses. A genome-wide association study identified two single-nucleotide polymorphisms in the exonic regions of the *SEMI-DWARF 37* ortholog that corresponded to the different *T. urartu* ecotype groups. Our study provides novel insights into the origin and genetic legacy of the A subgenome in polyploid wheat and contributes a gene repertoire for genomics-enabled improvements in wheat breeding.

Keywords: *Triticum urartu*, whole-genome resequencing, demographic history, geographic adaptation, selective sweep, GWAS

Wang X., Hu Y., He W., Yu K., Zhang C., Li Y., Yang W., Sun J., Li X., Zheng F., Zhou S., Kong L., Ling H., Zhao S., Liu D., and Zhang A. (2022). Whole-genome resequencing of the wheat A subgenome progenitor *Triticum urartu* provides insights into its demographic history and geographic adaptation. *Plant Comm.* **3**, 100345.

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.

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is one of the most important crops worldwide, providing approximately 20% of the calories consumed globally (FAOSTAT 2018; <http://faostat.fao.org/>). However, the genetic diversity within wheat cultivars has been drastically reduced during the process of domestication and breeding (Reif et al., 2005; Peng et al., 2011). Crop progenitors and their relatives represent a large gene repository for the improvement of desirable traits (Dempewolf et al., 2017). Thus, genomic characterization of wheat progenitors and identification of genes underlying phenotypic variations will provide important resources for wheat breeding (Rasheed and Xia, 2019).

As the donor of the A subgenome in tetraploid and hexaploid wheat, *T. urartu*, originating from the Fertile Crescent region, is crucial for understanding the evolution, domestication, and genetic improvement of bread wheat (Dvorák et al., 1993; Gustafson et al., 2009; Ling et al., 2013). The natural population of *T. urartu* exhibits a high level of phenotypic plasticity, e.g., in plant architecture, heading date, grain size, and other agronomic traits (Wang et al., 2017). Furthermore, *T. urartu* populations harbor beneficial alleles for resistance to biotic and abiotic stresses and tolerance to diverse environments (Roy et al., 2011; Huang et al., 2016) that are likely to be the consequence of natural selection and adaptation to diverse habitats in the Middle East (Brunazzi et al., 2018). Recently, a series of genes related to powdery mildew resistance (Qiu et al., 2005), stem rust resistance (Rouse and Jin, 2011), starch synthesis (Song et al., 2020), and seed storage proteins (Zhang et al., 2015; Shen et al., 2020; Luo et al., 2015, 2021) were characterized in *T. urartu* and subsequently deployed in cultivated wheat (Qiu et al., 2005; Song et al., 2020). For example, introgression of the Glu-A1 locus from *T. urartu* into durum wheat could enhance dough strength in their hybrid amphiploids (Alvarez et al., 2009). However, to facilitate more efficient utilization of these wild progenitors, it is necessary to unravel their genomic variations and population dynamics and further elucidate the genetic basis of geographic adaptation in *T. urartu*.

Whole-genome sequencing (WGS) has been used to identify genes responsible for geographic adaptation in crop plants such as rice (Meyer et al., 2016), maize (Romero Navarro et al., 2017), and barley (Russell et al., 2016). Moreover, major achievements have been made in decoding genomes of wheat and its progenitors, including those of bread wheat (International Wheat Genome Sequencing Consortium, 2018), wild emmer wheat (Avni et al., 2017), and *Aegilops tauschii* (Jia et al., 2013; Luo et al., 2017). The release of the *T. urartu* genome sequence (Ling et al., 2018) will provide effective tools for a wide range of functional and evolutionary studies that can contribute to its exploitation in wheat breeding.

In this study, 59 representative accessions of *T. urartu* were sequenced to identify genetic variations and population dynamics, including population divergence, differentiation, and gene introgression. The genomic regions associated with important agronomic traits, e.g., heading date, plant height, and grain length, were characterized. Signals conferring geographic adaptation were also detected, which will broaden the genetic bases for breeding by design in wheat.

RESULTS

Population sequencing, variant calling, and genetic diversity

According to our previous study (Wang et al., 2017), a set of 59 representative *T. urartu* accessions were selected for WGS, comprising 91.07% of the allelic variations of the germplasm panel (minor allele frequency >0.02) and most of the origin areas (Figure 1; Supplemental Tables 1 and 2). A total of 2.81 trillion base pairs of sequences were generated with a mean depth of 9.17× and covering 92.26% of the *T. urartu* G1812 genome (Ling et al., 2018) (Supplemental Table 3). A final set of 42 220 661 single-nucleotide polymorphisms (SNPs) and 3 013 722 insertions and deletions (indels) (shorter than 30 bp) were filtered and retained after quality control (Figure 2A; Supplemental Figure 1; Supplemental Tables 4 and 5; refer to the methods section). The accuracy of the called SNPs was further evaluated as 96.67% by PCR and Sanger sequencing with 35 randomly selected SNPs in 12 individual accessions (Supplemental Table 6).

An additional 162 accessions were subjected to genotyping by sequencing (GBS), and a total of 68.56 Gb sequences were generated. As a result, the number of SNPs for each accession above ranged from 10 715 to 81 847 (Supplemental Table 7). To further explore the evolutionary relationships among different A subgenomes, four einkorn wheat accessions (one *T. monococcum* and three *T. boeoticum* accessions) were sequenced simultaneously, with a total of 267.30 Gb sequences obtained. All the paired-end reads of 59 *T. urartu* and four einkorn wheat accessions, together with those from the A subgenome of bread wheat (cv. Chinese Spring), were aligned to the *T. urartu* genome (Supplemental Table 8).

The distribution of SNPs and indels in the 59 resequenced *T. urartu* accessions varied greatly among chromosomes (Figure 2B; Supplemental Figure 2; Supplemental Tables 9 and 10). Of the 42.2 million high-quality SNPs, most (40 660 883 SNPs or 96.31%) were located in intergenic regions, whereas genic regions contained 718 335 SNPs (1.70%), with 478 956 (1.13%), 17 673 (0.04%), 26 993 (0.06%), and 194 713 (0.46%) SNPs located in introns, 5' untranslated regions, 3' untranslated regions, and coding sequences (CDSs), respectively. There were 80 079 synonymous and 115 436 nonsynonymous SNPs in the CDSs, and the diversities (π) of synonymous and nonsynonymous SNPs were 0.55×10^{-3} and 0.42×10^{-3} across all genes, resulting in a synonymous/nonsynonymous substitution ratio of 1.32. Both nonsynonymous and synonymous polymorphisms decreased as the allele frequency increased, indicating that the low frequency of nonsynonymous and synonymous SNPs predominantly derived from very rare mutations (Supplemental Figure 3; Supplemental Table 9). For indels, 13 863 (0.46%) were located in coding regions, of which 1716 (12.38%) were frameshift mutations, and 2 822 821 (93.67%) were situated in intergenic regions (Supplemental Figure 4; Supplemental Table 10).

Nucleotide diversity (π) and Watterson's estimator (θ_{WV}) were estimated to be 2.595×10^{-3} and 2.353×10^{-3} , respectively, revealing a higher level of genetic diversity in *T. urartu* than in the A subgenome of domesticated hexaploid landraces



Figure 1. Geographic distribution of *T. urartu* accessions used in this study.

Red dots represent the locations of the eastern Mediterranean coastal (EMC) population, and blue dots represent the locations of the Mesopotamian-Transcaucasian (MT) population.

($\pi = 1.263 \times 10^{-3}$) (Cheng et al., 2019). Linkage disequilibrium (LD) for the whole *T. urartu* panel was measured to be 90 kb when the squared correlation coefficient (r^2) between pairs of SNPs decayed to half of its maximum value ($r^2 = 0.16$) (Figure 2C). This level of LD was much higher than that of wild soybean (27 kb) (Zhou et al., 2015), wild rice (20 kb) (Xu et al., 2012), and wild maize (22 kb) (Hufford et al., 2012).

Phylogenetic reconstruction and population structure

Phylogenetic reconstruction of the 59 resequenced *T. urartu* accessions based on WGS data showed two major clades (Supplemental Figure 5), which were designated the eastern Mediterranean coastal (EMC) and Mesopotamian-Transcaucasian (MT) populations. Individuals in the MT population were mainly from Turkey, Iraq, Iran, Armenia, and northern Syria, whereas those in the EMC population were mostly derived from Lebanon, Jordan, and southwestern Syria (Figure 1). To explore the evolutionary relationships among different A subgenomes, another phylogenetic tree was also constructed using einkorn wheat and the A subgenome of bread wheat as the outgroups (Figure 3A). The phylogenetic tree clearly confirmed that the divergence of *T. urartu* from einkorn wheat occurred much earlier than its allopolyploidization into common wheat (Marcussen et al., 2014).

Principal component analysis (PCA) of *T. urartu* populations recapitulated the clear-cut genetic differentiation observed in the phylogenetic tree, and the first two axes accounted for 22.16% and 9.27% of the total genetic variance, respectively. PC1 clearly

distinguished EMC and MT accessions, and PC2 further separated these two populations into minor subpopulations from different sampling sites and latitudes (Figure 3B; Supplemental Table 2). Bayesian clustering also revealed a population division, with the highest ΔK between the EMC and MT populations at $K = 2$ (Figure 3C; Supplemental Figure 6). When $K = 3$, northwestern Syrian accessions were separated from the MT population, and Jordanian accessions separated from the rest of the EMC population at $K = 4$. As K continued to increase, minor subpopulations from other regions such as Armenia, southern Syria, Turkey, Iran, and Iraq gradually separated.

Two genetic diversity parameters, π and θ_W , were calculated in different *T. urartu* populations (Table 1). The MT population ($\pi = 1.654 \times 10^{-3}$, $\theta_W = 1.800 \times 10^{-3}$) possessed higher diversity than the EMC population ($\pi = 1.103 \times 10^{-3}$, $\theta_W = 1.120 \times 10^{-3}$). The estimated population-differentiation statistic (F_{ST}) between the two populations was 0.310, indicating a moderate level of differentiation within the whole *T. urartu* population. This result is consistent with their geographic distributions and discrete clusters in the PCA space (Figures 1 and 3B). In addition, significant differences in LD block number and density across chromosomes were also observed between the MT and EMC populations, probably due to the large variations in SNP distribution and position in each chromosome (Figure 3D; Supplemental Figure 7; Supplemental Table 11). There were 208 791 and 180 575 LD blocks identified in the MT and EMC populations, respectively, accounting for 3 695 383 and 3 218 423 kb in length (Supplemental Table 12). Consequently, LD

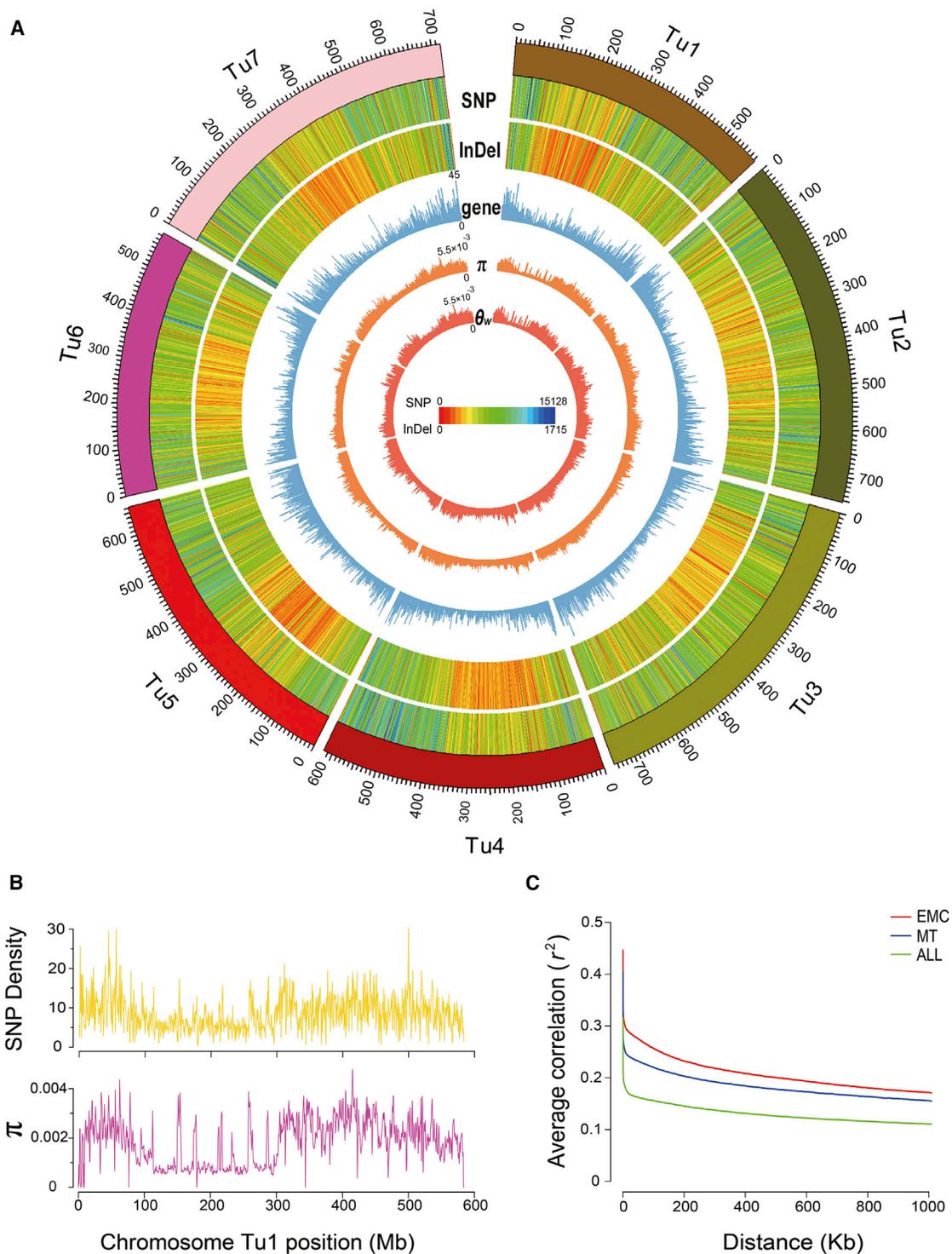


Figure 2. Genomic variation map of *T. urartu*.

(A) Circos plot showing different features of the *T. urartu* genome. The chromosomes are numbered. Tracks from outer to inner are the seven chromosomes (one scale label indicates 10 Mb), SNP density, insertion or deletion density, gene density (with a window size of 500 kb and a step size of 100 kb), nucleotide diversity (π), and Watterson's theta (θ_w).

(B) Close-up view of chromosome Tu1 showing SNP density and π in nonoverlapping 100-kb bins. The yellow line represents SNP density, and the purple line represents π . For other chromosomes, see [Supplemental Figure 2](#).

(C) Decay of linkage disequilibrium in the whole *T. urartu* panel and the two subpopulations measured by the squared correlation coefficient (r^2) against distance.

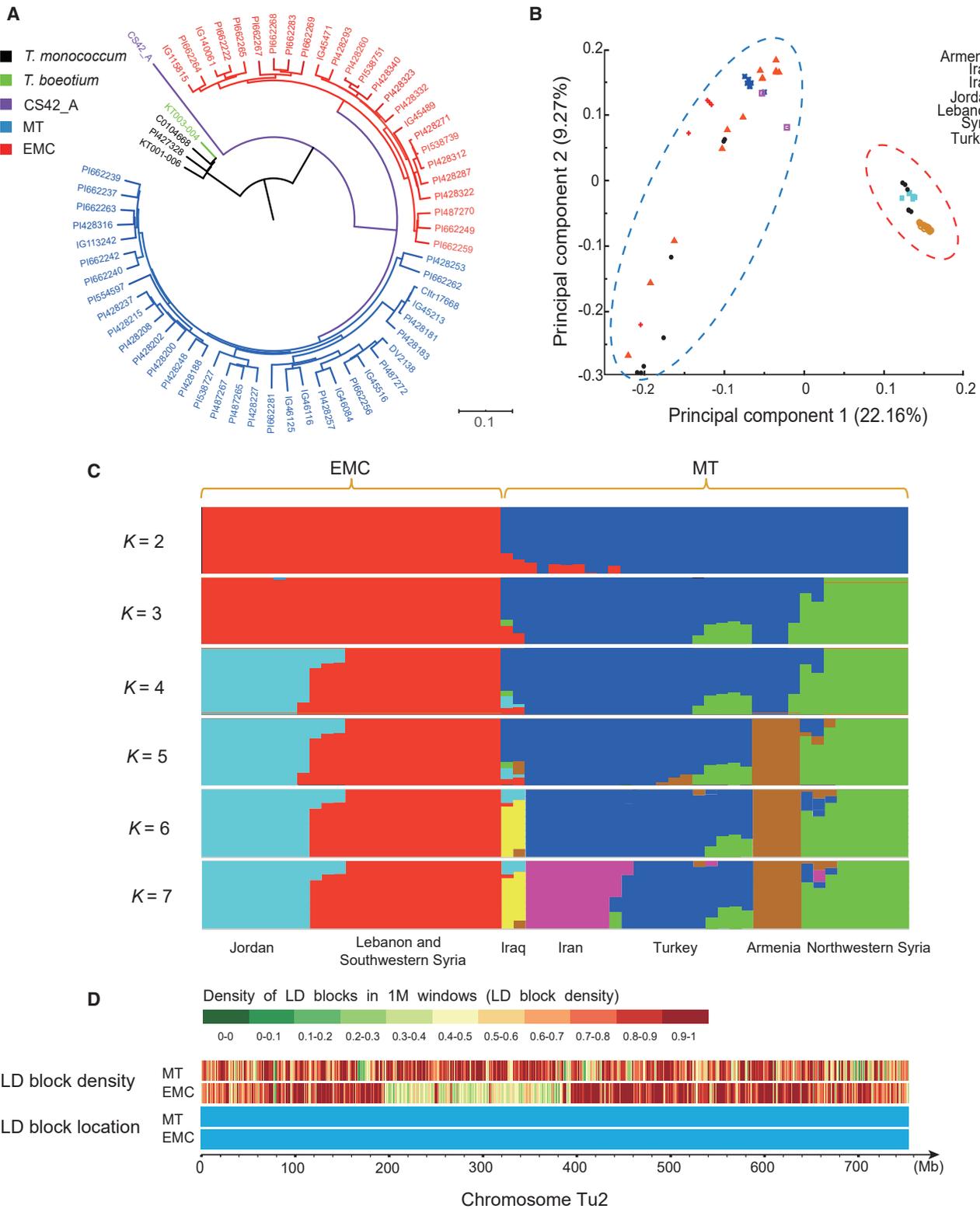


Figure 3. Population structure and haplotype block analysis of *T. urartu* accessions.

(A) A phylogenetic tree of 59 *T. urartu* accessions constructed based on the whole-genome SNP analysis using einkorn wheat as an outgroup. The tree clearly reveals that all the *T. urartu* accessions can be divided into EMC and MT groups and that *T. urartu* diverged from einkorn wheat far earlier than its hybridization into polyploid wheat. CS42-A, A subgenome of Chinese Spring.

(B) Principal component analysis of the first two components. The first and second principal components account for 22.16% and 9.27% of the total variation, respectively.

(legend continued on next page)

Group	π (10^{-3})	θ_w (10^{-3})	Tajima's <i>D</i>	F_{ST}
EMC ^a	1.103	1.120	0.207	–
MT ^b	1.654	1.800	0.183	–
EMC/MT	0.667	0.622	1.131	–
All samples	2.595	2.353	–0.054	0.310

Table 1. Statistics of diversity levels in different *Triticum urartu* populations.

^aEMC, eastern Mediterranean coastal.

^bMT, Mesopotamian-Transcaucasian.

decayed to its half maximum value within 236 kb in the MT group ($r^2 = 0.20$) and 275 kb in the EMC group ($r^2 = 0.22$) (Figure 2C).

Evolutionary and demographic history

Investigating gene flow between different geographic origin groups of *T. urartu* could provide an opportunity to understand the evolutionary process. The divergence and migration events among these groups were therefore examined based on a Gaussian model of allele frequency change (Pickrell and Pritchard, 2012). In light of the geographic distribution of origin areas and the potential genetic boundaries among subpopulations, 59 *T. urartu* accessions were divided into seven regional groups (Supplemental Figure 8; Supplemental Table 13). A maximum-likelihood (ML) tree was generated using WGS data, and a primary split between western and eastern wings of the Fertile Crescent (representing the EMC and MT populations, respectively) was observed (Supplemental Figure 9A). Certain patterns of the topology persisted and were robust even if the number of migration events (m) was permitted to change within the model. Allowing just one migration event ($m = 1$), we observed an introgression from the Jordan into the Turkey group with a migration weight of 14.07% (Supplemental Figure 9B). Allowing two migration events ($m = 2$), a substantial amount of gene flow (37.87%) was detected from the Turkey to the northwestern Syria group (both from the MT population) (Supplemental Figure 9C). When $m = 3$, a gene flow (7.39%) was inferred from the Armenia to the Iraq group along the abovementioned migration events (Figure 4A). Without migration events, the TreeMix ML tree could explain 98.51% of the variance in relatedness among groups, and inclusion of the first three migration events increased the explained variance to 99.94%. These analyses can offer an overview of gene flow among different *T. urartu* geographic accessions.

In particular, the Armenia group had a long branch in the topology, suggesting a strong bottleneck, probably due to specific geographic adaptation (i.e., during a glacial stage approximately 12 900 to 11 700 years before the present, named the Younger Dryas) (Jones et al., 1998; Haldorsen et al., 2011). This group was located in the Caucasian region, a domestication center from which bread wheat originated (Charmet, 2011). It seemed

that *T. urartu* accessions from Armenia spread into Iraq, accompanied by bread wheat cultivation. Likewise, the northwestern Syria group also possessed an exceptionally long branch and more genetic drift in topology, similar to the Armenia group (Figure 4A). Interestingly, when the A subgenome of Chinese Spring (CS42) was used as an outgroup, a significant genetic drift from the northwestern Syria group (around the Karacadag mountains) to hexaploid wheat was observed to contribute 42.85% of the DNA (Figure 4B), which was coincident with the origin center of tetraploid wheat (Charmet, 2011; Peng et al., 2011; Abbo and Gopher, 2017; Zhou et al., 2020). Thus, *T. urartu* derived from this area was probably one of the potential progenitors of both tetraploid and hexaploid wheat.

Historical changes in effective population size (N_e) in *T. urartu* were reconstructed using the sequential Markov coalescent (SMC++) model (Terhorst et al., 2017) with a mutation rate of 5.5×10^{-9} per nucleotide per year (Dvorak and Akhunov, 2005; Chao et al., 2009). We found that the ancient *T. urartu* population experienced a very large reduction in N_e from $\sim 3\,000\,000$ to a plateau of $\sim 140\,000$ and subsequently split into EMC and MT clades approximately 12.9 kilo years ago (kya) during the famous Younger Dryas episode (Renssen et al., 2015) (Figure 4C). Previous studies have shown that the Fertile Crescent regions underwent extreme climate change during this period. For instance, the late Pleistocene climate of the EMC was characterized as cold and dry, whereas the inland Near East around the Karacadag mountains was a relatively moist area and is thought to have been a refugia for wheat and many other plants (Wright, 1976; Zeder, 2011). Thus, the potentially abrupt historical event between the EMC and MT populations possibly brought about their divergence. Similarly, the SMC++ profiles also indicated a reduction in N_e for both EMC and MT populations that commenced 12.8–13.0 kya and lasted until 1.3–1.5 kya. Unlike that of the EMC population, however, the N_e of the MT population briefly recovered from its nadir approximately 1.4–3.2 kya. Archaeological evidence has demonstrated that eastern Fertile Crescent regions near the Tigris and Euphrates rivers were going through the late Bronze Age (~ 1400 to 1200 BC) during the same period, during which human activities such as settlement, grazing, and agriculture became major drivers,

(C) Population structure plots with different numbers of clusters ($K = 2-7$). A major division was detected between the EMC and MT populations (represented by red and dark blue colors, respectively) at $K = 2$. When $K = 3$, northwestern Syrian accessions (green) were separated from the MT population, and Jordanian accessions (light blue) were separated from the rest of the EMC population when $K = 4$. As K continued to increase, minor subpopulations from other regions such as Armenia (brown), Iraq (yellow), and Iran (purple) gradually separated.

(D) Distribution of haplotype blocks on chromosome Tu2 for the two different elite populations. Color intensity conveys the extent of linkage between pairs of SNP markers (red, highest). For other chromosomes, see Supplemental Figure 7.

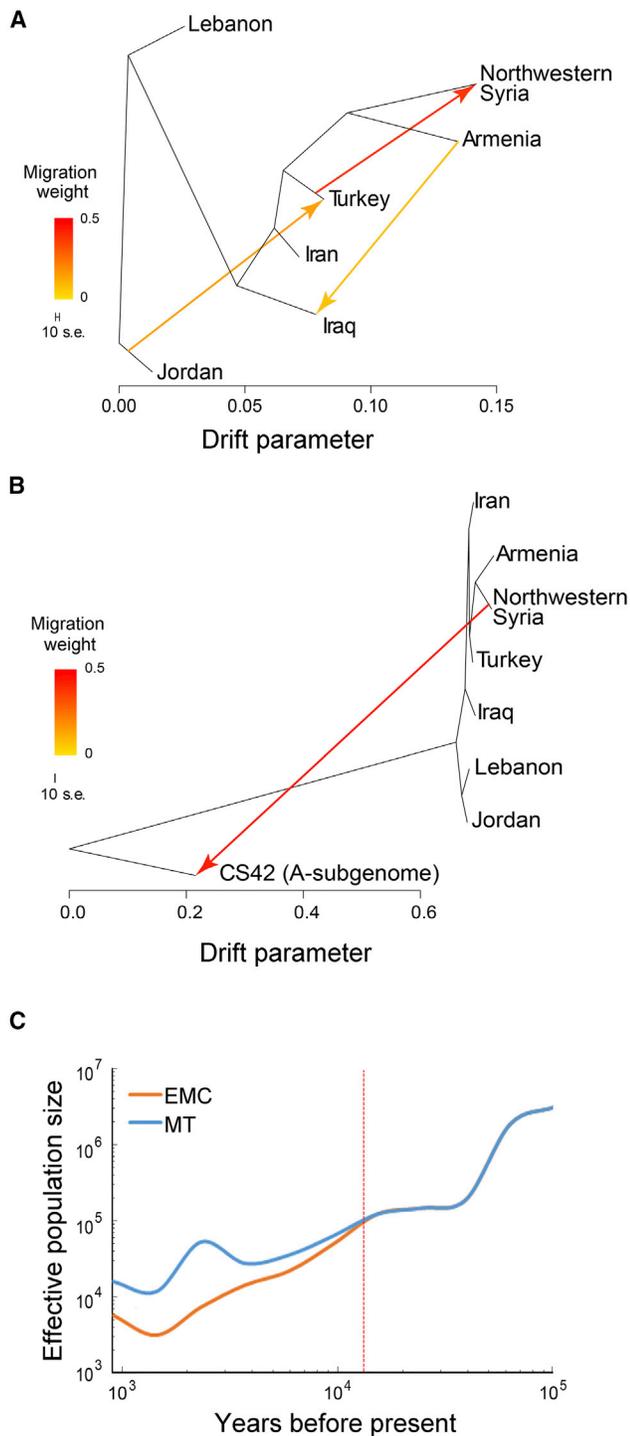


Figure 4. Demography of *T. urartu*.

(A) TreeMix analysis of 59 *T. urartu* samples divided into seven geographic groups (Supplemental Figure 8). Arrows represent migration events among different groups. The weight of the migration component is scaled according to the color scale on the left.

(B) TreeMix analysis of *T. urartu* samples using the A subgenome of Chinese Spring as an outgroup. The arrow of maximum-likelihood topology indicates that *T. urartu* accessions from northwestern Syria have the closest consanguinity with the A subgenome of Chinese Spring. This finding indicates that tetraploid wheat may have originated from northwestern Syria (around the Karacadag mountains).

accounting for changes in landscape (Knapp, 1992; Klinge and Fall, 2010; MCGovern, 2010). This may be the reason why the MT population had more polymorphisms and a wider distribution than the EMC population, consistent with its expansion of effective population size. In addition, the divergence between the direct donor of the wheat A subgenome and *T. urartu* was estimated to have occurred approximately 0.42 million years ago using SMC++ analysis, consistent with previous studies (Ling et al., 2013; Baidouri et al., 2017) (Supplemental Figure 10).

Geographic adaptation in the MT population

Changes in allele frequency have been utilized for identifying footprints of geographic adaptation during evolution. Through F_{ST} and nucleotide diversity ratio ($\ln \pi_{EMC}/\pi_{MT}$) analysis between the MT and EMC populations (Figure 5A), genomic regions were characterized with strong genetic differentiation using a top 5% cutoff and were considered to be putative sweep-selected sections (Figure 5B). These genomic regions comprised approximately 2.31% (112.10 Mb) of the genome and were mainly located on the long-arm region of chromosome Tu2. Compared with the EMC population, the MT population harbored significantly lower nucleotide diversity in these genomic regions, resulting in a long LD fragment (Supplemental Figures 11 and 12; Supplemental Table 14). The characteristics of these regions were also confirmed by significantly lower values of Tajima's D ($p = 2.20 \times 10^{-16}$, Wilcoxon rank-sum test) and extended LD ($p = 2.20 \times 10^{-16}$, Wilcoxon rank-sum test) in the MT population (Supplemental Figure 13), in agreement with the distribution of haplotype blocks on chromosome Tu2 (Figure 3D; Supplemental Figure 7). Regions under sweep selection were also screened with the XP-CLR and XP-EHH methods (Sabeti et al., 2007; Chen et al., 2010), resulting in 229.20 (858 genes) and 170.10 Mb (1,941 genes) of putatively selected regions, respectively (Figure 5C and 5D; Supplemental Tables 15 and 16).

In summary, a total of 2844 genes were identified from the above three approaches, providing potential targets of geographic adaptation in the MT population (Figure 5E; Supplemental Table 17). Gene Ontology (GO) and KEGG enrichment analyses were performed to explore the biological functions of these candidate genes, and several highly enriched functional categories were identified, including cell differentiation (GO:0030154) and photosynthesis (GO:0009523, ko00195) (Supplemental Figures 14 and 15; Supplemental Tables 18 and 19). These genes, especially those related to plant development and photosynthesis, contribute to the divergence of the MT and EMC populations.

Among these candidate genes, TuG1812G0200002519 is a homolog of *HD16* located in the centromeric region of chromosome Tu2 and encodes a casein kinase I protein that phosphorylates downstream genes (such as *FT1*) in the flowering pathway (Kwon et al., 2015) (Supplemental Table 20). *HD16* plays an important role in controlling heading date in rice and contributes to rice growth in a wide range of climates (Hori

(C) Split time when EMC and MT populations separated from their ancestry inferred by SMC++. The dashed line indicates the split time nearly 12 900 years ago.

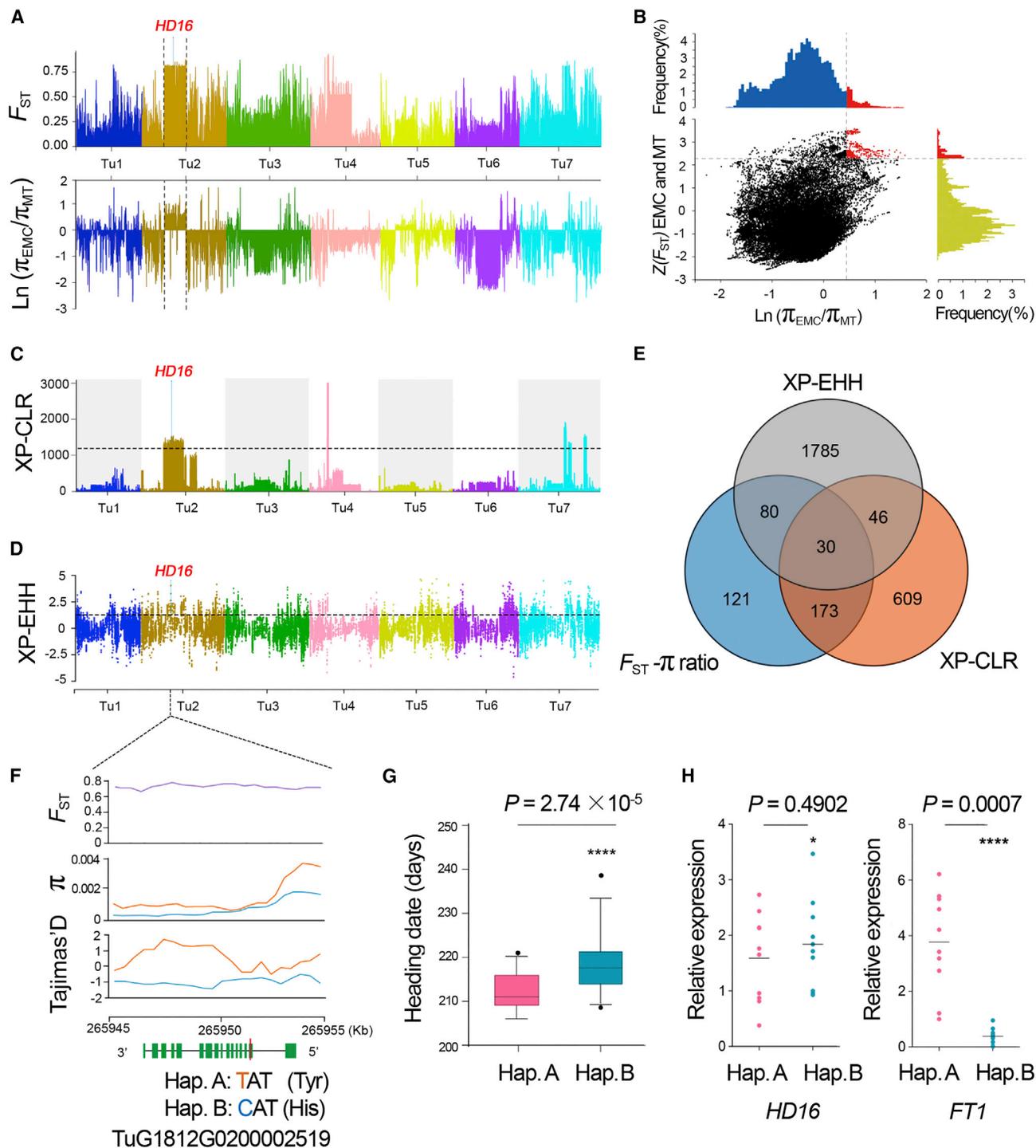


Figure 5. Genomic regions with geographic selective sweeps in the MT population.

(A) The distribution of F_{ST} and $\ln \pi$ -ratio (EMC/MT) along chromosomes. The genome-wide threshold is defined by the top 5% of the F_{ST} and $\ln \pi$ -ratio (EMC/MT) values. Candidate genes are marked on the map.

(B) Intersection of $\ln \pi$ -ratio (EMC/MT) and $Z(F_{ST})$ values calculated in 500-kb sliding windows with 100-kb steps. Red dots represent genomic regions under geographic adaptation in the MT population (corresponding to $\ln \pi$ ratios >0.44 and $Z(F_{ST}) > 2.35$).

(C) XP-CLR values plotted against the position on each of the seven chromosomes. The horizontal dashed lines indicate the genome-wide threshold for selection signals. The adaptive genes identified above are mapped to genomic regions with high XP-CLR values.

(D) XP-EHH values plotted against the position on each of the seven chromosomes. The horizontal dashed lines indicate the genome-wide threshold for selection signals. The adaptive genes identified above are mapped to genomic regions with high XP-EHH values.

(E) Venn diagram of genes under selection that were detected using the $F_{ST}-\pi$ ratio, XP-CLR, and XP-EHH methods.

(legend continued on next page)

et al., 2013). In our study, this gene had a low level of π in MT and a high level of F_{ST} between the two populations (Figure 5F), and a SNP in the conserved casein kinase domain introduced a tyrosine-to-histidine substitution. Haplotype analysis also demonstrated a clear separation between MT and EMC, in which Hap A (TAT, Tyr124) and B (CAT, His124) were found mainly in early- and late-heading ecotypes, respectively (Figure 5G). The expression levels of *HD16* and *FT1* orthologs in the leaves of Hap. A and Hap. B accessions were compared through quantitative real-time PCR analysis. Transcript levels of the *HD16* ortholog (TuG1812G0200002519) were higher in Hap. B accessions than in Hap. A accessions. By contrast, the *FT1* ortholog (TuG1812G0300001662) showed significantly higher expression in Hap. A than in Hap. B accessions (Figure 5H). It seemed that the *HD16* ortholog with ecotype-specific SNPs negatively regulated *FT1* transcription levels, resulting in differences in flowering time between the two populations, consistent with the findings of previous studies (Hori et al., 2013).

Geographic adaptation in the EMC population

To examine the adaptive genome patterns of the EMC population, genomic regions with extreme allele frequency differentiation were scanned with various methods. In the top 5% of F_{ST} regions, no significant π -ratio signal ($\ln \pi_{EMC}/\pi_{MT}$) was observed in the EMC population. However, XP-CLR and XP-EHH led to the identification of 476 (1674 genes, 288.20 Mb) and 272 (1980 genes, 168.44 Mb) putatively selected regions, respectively (Figure 6A and 6B; Supplemental Tables 21 and 22). Overall, a total of 3505 candidate adaptation-related genes were detected in the EMC population, and we next tested whether these genes were associated with specific biological processes or annotation categories by exploratory functional enrichment analysis (Figure 6C; Supplemental Table 23). Using Fisher's exact tests, 37 significantly enriched GO terms ($p < 0.05$) were identified in the three major GO categories: oxidation–reduction process, defense response, and eight other terms in the biological process category; membrane, cell periphery, and nine other terms in the cellular component category; and monooxygenase activity, oxidoreductase activity, and fourteen other terms in the molecular function category (Supplemental Figure 16; Supplemental Table 24). These 3505 genes were also enriched in 14 KEGG pathways ($p < 0.05$), especially the plant hormone signal transduction, steroid biosynthesis, and plant-pathogen interaction pathways (Supplemental Figure 17; Supplemental Table 25).

All plant species survive and propagate under natural conditions, and environmental factor–related traits, such as photoperiod, dormancy, and stress response, contribute to their adaptation characteristics. The candidate genes responsible for these traits, identified through XP-CLR and XP-EHH approaches, have been verified in wheat, rice, and other plants (Figure 6A and 6B; Supplemental Table 20). For example, several polycomb and

phytochrome groups clustered in the terminal regions of chromosome 2AS, including four genes characterized as *ETR2* (TuG1812G0200000032) (Dahleen et al., 2012), *EMF2* (TuG1812G0200000033) (Chen et al., 2009), *CYP96B4* (TuG1812G0200000045) (Rengasamy et al., 2011; Zhang et al., 2014; Tamiru et al., 2015), and *DWARF* (TuG1812G0200000103) (Hong et al., 2002). Phytochromes, such as cytokinin and ethylene receptors, are major mediators of red and far-red light responses that play vital roles in plant growth, development, and reproduction (Wang and Wang, 2015). Similarly, the cryptochrome gene *CRY1a* (TuG1812G0600002035), mapped to chromosome 6A, encodes blue-light receptors that induce downstream light responses such as photomorphogenesis, enhancement of cotyledon expansion, and osmotic stress tolerance in wheat (Xu et al., 2009). In addition, one photoperiodic gene *FTIP1* (TuG1812G0700004364), previously reported to control *FT* (*VRN-3*) movement to the shoot apical meristem and influence flowering time (Song et al., 2017), was identified on chromosome 7AL.

Seed dormancy and its release is another important adaptive trait in plants, largely shaped by photoperiod, hormones, and temperature regimes (Evans et al., 2014). In our study, three conserved orthologs that emerged in the XP-CLR and XP-EHH outliers, namely, *HUB1* (TuG1812G0300004067) (Cao et al., 2015), *FIE2* (TuG1812G0700003420) (Dickinson et al., 2012; Nallamilli et al., 2013), and *PTR2* (TuG1812G0700004450) (Fang et al., 2013), have been shown to be the major regulators of seed dormancy in *Arabidopsis*, rice, and wheat. In addition, a large number of beneficial genes that confer resistance to various abiotic and biotic stresses, particularly drought, salinity, and pathogen infection, were overrepresented in the putative sweep regions, including *HAK21* (TuG1812G0200003042) (Zhang et al., 2012), *SAMS1* (TuG1812G0300002441) (Bhuiyan et al., 2007; Kasim et al., 2013), *SGT1* (TuG1812G0300002657) (Wang et al., 2015), *BIHD1* (TuG1812G0400000233) (Luo et al., 2005), and *MPK13* (TuG1812G0700004443) (Lian et al., 2012). Genes affecting plant architecture, such as *LSK1* (TuG1812G0200003175) (Zou et al., 2015) and *PAY1* (TuG1812G0500000051) (Zhao et al., 2015), were also detected. The exploration of these gene alleles can provide numerous genetic resources for wheat breeding.

To further annotate the adaptive genomic regions, we investigated 10 agronomic and grain traits of 221 *T. urartu* accessions in three environments for two consecutive years (2013–2015) (Wang et al., 2017) (Supplemental Tables 26–28). The panel of analyzed *T. urartu* accessions exhibited high phenotypic diversity, with substantial variation in most traits. Specifically, the Armenian accessions had an extremely late heading date, a higher number of spikelets per spike, and smaller grains than other accessions, consistent with their genetically distinct position and greater genetic drift in the TreeMix tree (Figure 4A; Supplemental Figure 18). We also combined the WGS and GBS data and

(F) Example of a gene (TuG1812G0200002519) with selective sweep signals in the MT population. F_{ST} , π , and Tajima's D values are plotted using a 1-kb sliding window. Genes are shown at the bottom (exons and introns are represented by boxes and lines, respectively). The position of the causal SNP is marked by a red line. Hap., haplotype.

(G) Comparisons of heading date (day) between haplotypes A and B in the whole-genome resequenced *T. urartu* population. In boxplots, center line indicates median; box limits indicate upper and lower quartiles; whiskers denote 1.5 \times interquartile range; and points show outliers.

(H) Expression levels of the *HD16* ortholog (TuG1812G0200002519) and *FT1* ortholog (TuG1812G0300001662) in different populations revealed by quantitative real-time PCR analysis. Differences between two haplotypes were analyzed with an unpaired t test.

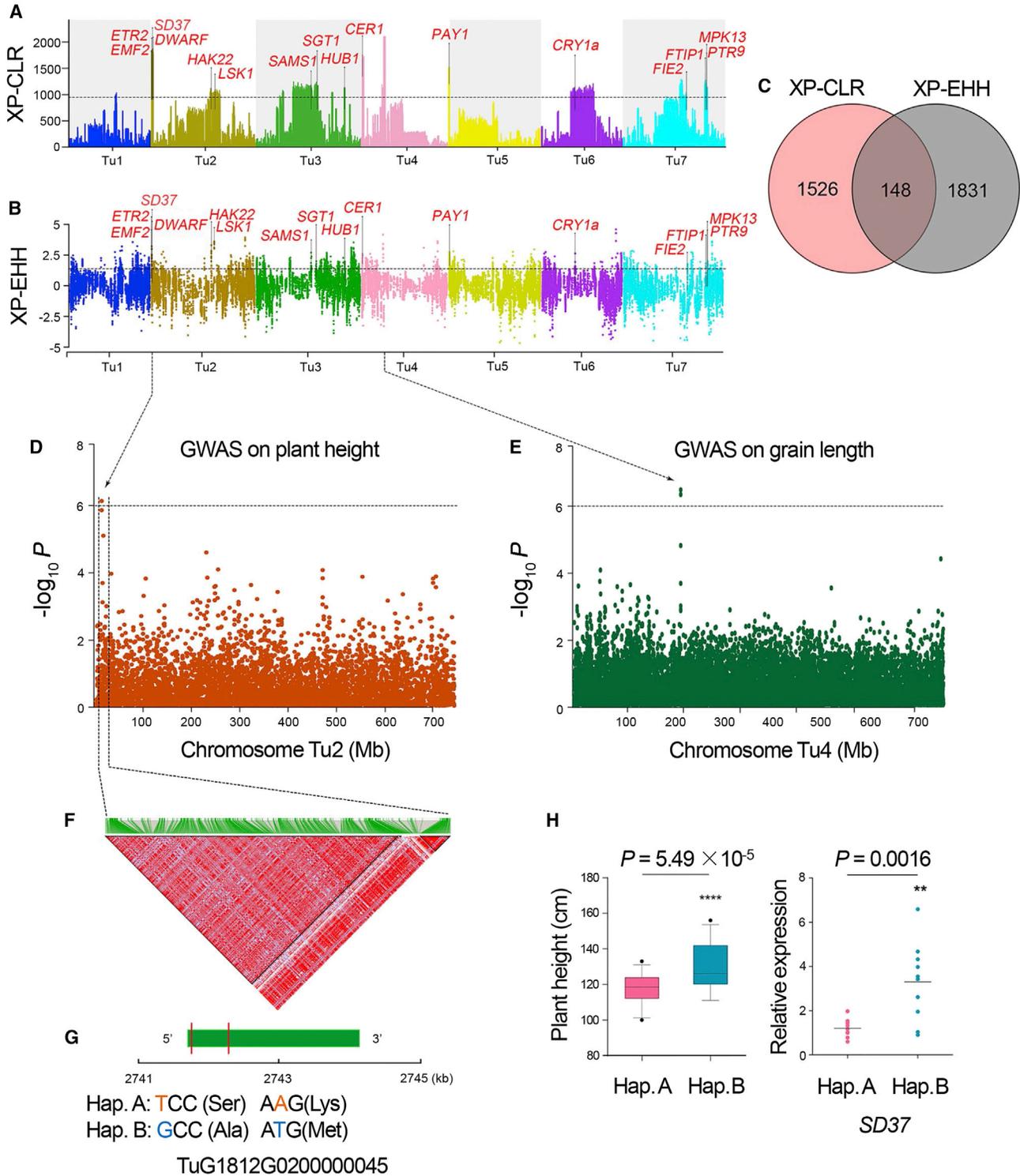


Figure 6. Genome-wide screening and functional annotations of selected regions during geographic adaptation in the EMC population.

(A) XP-CLR values plotted against the position on each of the seven chromosomes. The horizontal dashed lines indicate the genome-wide threshold for selection signals. The adaptive genes identified above are mapped to genomic regions with high XP-CLR values.

(B) XP-EHH values plotted against the position on each of the seven chromosomes. The horizontal dashed lines indicate the genome-wide threshold for selection signals. The adaptive genes identified above are mapped to genomic regions with high XP-EHH values.

(C) Venn diagram of genes under selection detected using the XP-CLR and XP-EHH methods.

(D and E) Genome-wide association study results for plant height and grain length that overlapped by strong selective signals. Dashed horizontal lines indicate the significance thresholds ($-\log_{10} p = 5.92$ in D, $-\log_{10} p = 6.34$ in E).

(legend continued on next page)

obtained 149,030 high-quality SNPs after filtering based on minor allele frequency (MAF) >0.05. Afterwards, genome-wide association study (GWAS) analyses were performed using TASSEL 5.0 software via a compressed mixed linear model (MLM) (Bradbury et al., 2007). The p-value thresholds were adjusted by permutation-based false discovery rate (FDR) to identify significant loci. Finally, a GWAS signal associated with plant height was detected on chromosome 2AS (chrTu2: 2,752,902), which was located in the XP-CLR and XP-EHH outlier regions and coincided with quantitative trait loci (QTLs) responsible for shorter stature and lodging resistance reported in spelt wheat (Keller et al., 1999) (Figure 6D; Supplemental Figure 19A; Supplemental Table 29). Further LD analysis of this region (2.74–2.75 Mb; Figure 6F) revealed that an ortholog of cytochrome P450 monooxygenase *CYP96B4* or *SD37* (TuG1812G0200000045) was predicted to underlie this QTL for plant height, as two causative exonic SNPs were detected in the strong XP-CLR and XP-EHH outliers (Figure 6G). Haplotype analysis showed that Hap. A (TCC, Ser28 and AAG, Lys386) was mainly detected in short accessions from low latitudes, whereas Hap. B (GCC, Ala28 and ATG, Met386) was observed in tall accessions, which tended to occur at high latitudes (Figure 6H). Previous studies have shown that *CYP96B4* is one of the semidwarf genes in rice that triggered the “green revolution” and widely enhanced the harvest index and biomass production of crops (Rengasamy et al., 2011; Zhang et al., 2014; Tamiru et al., 2015). Our results showed that the expression levels of *CYP96B4* (TuG1812G0200000045) in stems differed significantly between the two haplotype accessions (Figure 6H), supporting the idea that there may be a switch in plant strategy between height and latitude (Moles et al., 2009). In addition, we also detected another locus on chromosome 4AS that was associated with grain length (Figure 6E), with a linked LD extent of 100 kb overlapping with the adaptive genomic regions (Supplemental Figure 19B; Supplemental Table 30). These QTLs could provide insights into factors that are crucial for the adaptation of wheat and its relatives to specific environmental conditions (Maccaferri et al., 2008).

DISCUSSION

As one of its wild relatives, *T. urartu* plays a central role in the development and evolution of cultivated bread wheat (Dvorák et al., 1993; Gustafson et al., 2009; Ling et al., 2013). In this study, the comprehensive landscape of polymorphisms in *T. urartu* provides novel insights into the population characteristics, evolutionary history, and geographic adaptation of the A subgenome progenitor and, finally, contributes to the identification of a series of genomic regions and genes for important traits and stress resistance.

The discovery of genome-wide patterns of genetic variation within *T. urartu* will benefit future marker-assisted mapping and guide breeding strategies for the behavior of the wheat A subgenome.

In this study, more than 40 million high-quality SNPs and 3 million indels were obtained from 59 *T. urartu* representatives through the WGS strategy. The genetic diversity of *T. urartu* was lower than that of other wild crops, e.g., wild soybean ($\pi = 2.94 \times 10^{-3}$) (Zhou et al., 2015), wild rice ($\pi = 3.34 \times 10^{-3}$) (Xu et al., 2012), and wild barley ($\pi = 2.97 \times 10^{-3}$) (Pankin et al., 2018), but it was much higher than that of their cultivated relatives ($\pi = 1.05 \times 10^{-3}$, $\pi = 2.14 \times 10^{-3}$, and $\pi = 1.53 \times 10^{-3}$, respectively). However, both wild and cultivated maize retained greater nucleotide diversity than *T. urartu*, probably due to their open pollination habit and hybrid varieties ($\pi = 5.90 \times 10^{-3}$ and 4.80×10^{-3}) (Hufford et al., 2012). These findings indicated that our results were credible and reasonable, as *T. urartu* is a mainly self-pollinating wild species. On this basis, a detailed LD block map of the A subgenome was generated, which contains highly differentiated regions that are crucial for selection and geographic adaptation. In addition, the relatively rapid decay of LD in *T. urartu* suggests that GWAS should enable high-resolution mapping of genes associated with natural phenotypic variation.

Information from both genome-wide SNP data and simple-sequence-repeat markers clearly assigned *T. urartu* accessions to the EMC and MT populations (Hammer et al., 2000; Wang et al., 2017). Through F_{ST} and π ratios, XP-CLR, and XP-EHH analyses along the seven chromosomes, we successfully characterized genomic regions with unusually long haplotypes that are related to genetic differentiation between the two major groups; these were also confirmed by significantly lower Tajima's D and extended LD in the MT population. Our results showed that $F_{ST-\theta_\pi}$ and XP-CLR detected the most shared genes, whereas XP-CLR and XP-EHH detected the lowest number of shared genes. Combining different statistical methods can minimize biases and false positives and is therefore considered an optimal strategy for detecting selection signatures (Vatsiou et al., 2016). Genes located in these selective-sweep regions were involved in specific functions such as cell differentiation (GO:0030154) and photosynthesis (GO:0009523, ko00195), and they may have contributed to the divergence of these two populations. Specifically, one candidate for geographic adaptation homologous to *HD16* was identified, which was demonstrated to negatively regulate flowering time in rice (Hori et al., 2013). In our study, haplotype and expression analyses of the *HD16* ortholog in *T. urartu* showed that its ecotype-specific SNPs may result in differences in flowering time between the two populations. However, given the wide nonrecombining region on chromosome Tu2, there may also be other important candidate genes in this large interval that contribute to the difference in flowering time between MT and EMC populations (Adamski et al., 2020). On the other hand, the gene flow within and between subgroups of different areas demonstrated that northwestern Syrian and Armenian accessions were genetically distinct and showed more genetic drift, consistent with archaeological data showing that tetraploid and hexaploid wheat originated from these

(F) Linkage disequilibrium heatmap of the ~100-kb *SD37* region (2.74–2.75 Mb).

(G) Zoomed-in view of the candidate gene (TuG1812G0200000045) associated with SNP polymorphism.

(H) Comparisons of plant height (cm) between haplotypes A and B in the whole-genome resequenced *T. urartu* population. In boxplots, center line indicates median; box limits indicate upper and lower quartiles; whiskers denote 1.5× interquartile range; and points show outliers. Expression levels of the *SD37* ortholog (TuG1812G0200000045) in different haplotype accessions were revealed by quantitative real-time PCR analysis. Differences between two haplotypes were analyzed with an unpaired t test.

Plant Communications

two regions, respectively (Charmet, 2011). Furthermore, the accessions from the northwestern Syrian subgroup were predicted to be the progenitor of the A subgenome of tetraploid and hexaploid wheat. Apart from these, our SMC++ analysis also demonstrated that the direct donor of the A and B subgenomes hybridized to form tetraploid emmer wheat approximately 0.42 million years ago around northwestern Syria, whereas the ancestor of the *T. urartu* A subgenome proceeded to evolve naturally and subsequently split into EMC and MT populations during the Younger Dryas period. Previous studies have shown that the Younger Dryas event was a boundary for wheat domestication, during which the Fertile Crescent regions underwent extreme climate change (Jones et al., 1998; Haldorsen et al., 2011). These results indicate that the MT population may have undergone specific geographic adaptation during the process of wheat domestication (Abbo and Gopher, 2017).

Native species that are subjected to a diverse range of environments over long periods possess valuable resources for germplasm improvement. In our study, we scanned the A subgenome for selection signals using the XP-CLR and XP-EHH methods and identified 730 sweeps that might underlie geographic adaptation in the EMC population. These regions contained a large number of phenological and photoperiodic genes, including those that encode the photosynthetic apparatus (e.g., *ETR2* and *CRY1a*) (Xu et al., 2009; Dahleen et al., 2012) adaptable to light quality changes and the *FT*-interacting protein (e.g., *FTIP1*) regulating flowering time (Song et al., 2017), which significantly influence growth capacity and reproduction. We also detected candidate adaptive genes that are related to seed dormancy (e.g., *HUB1*), which could help plants survive in harsh environments by delaying and desynchronizing germination (Cao et al., 2015). In addition, there are numerous beneficial genes or their alleles in the selection outliers that cope with abiotic and biotic stresses, such as the *HAK* potassium transporter (*HAK21*), S-adenosyl-methionine synthase (*SAMS1*), and main components of the R protein (*SGT1*). They are crucial for plant resistance to salinity, drought, and pathogen infection (Luo et al., 2005; Bhuiyan et al., 2007; Lian et al., 2012; Zhang et al., 2012; Wang et al., 2015). The above analysis will improve our understanding of the genetic basis of geographic adaptation in *T. urartu* and other wheat species.

Taken together, the catalog of polymorphisms in *T. urartu* can supply a foundational resource for marker-assisted selection and the discovery of candidate mutations associated with important agronomic traits. As evidenced in this study, short stature and lodging resistance in *T. urartu* and spelt wheat arose from the same QTL, suggesting that the A subgenome may share most of its genetic and molecular information to facilitate agronomic improvement. The genome-wide SNPs of *T. urartu*, together with the growing availability of other comparative genomic data, will unravel the nature of genetic and phenotypic diversity and accelerate tailor-designed breeding in common wheat.

METHODS

Sampling information

A total of 221 diverse *T. urartu* accessions were collected from the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences

Whole-genome resequencing of *Triticum urartu*

(Beijing, China), the National Small Grains Collection (USDA-ARS; Pullman, WA, USA), and the International Center for Agricultural Research in the Dry Areas (Beirut, Lebanon). A core germplasm collection of 59 representative accessions was used for whole-genome resequencing (Supplemental Table 1), and the remaining 162 accessions were subjected to GBS (Supplemental Table 6). In addition, four einkorn wheat materials were used as outgroups, including three *T. boeoticum* accessions (C0104668, PI427328, and KT001-006) and one *T. monococcum* accession (KT003-004) (Supplemental Table 8).

Library construction and sequencing

Genomic DNA was extracted from young leaf tissues of 2-week-old seedlings (one individual per accession) using the cetyl trimethyl ammonium bromide method as described elsewhere (Saghai-Marooof et al., 1984). WGS libraries with insert sizes of approximately 300 or 500 bp were constructed using a paired-end DNA sample preparation kit (Illumina, San Diego, CA, USA). With respect to GBS, library preparations were manipulated with the *MspI* and *BamHI* restriction enzymes following Elshire et al. (2011), and DNA fragments with insert sizes ranging from 300 to 600 bp were selected on Invitrogen E-Gel SizeSelect 2% (Invitrogen, Carlsbad, CA, USA). We sequenced 100 or 150 bp at each end using the HiSeq 2000 or HiSeq 4000 platform (Illumina).

Read alignment and variant calling

SOAPnuke software (<https://github.com/BGI-flexlab/SOAPnuke>) (Chen et al., 2018) was used to remove adaptor sequences and low-quality reads (defined as those with >50% bases with a quality score of <10). The reads were aligned to the reference sequence of *T. urartu* using the Burrows-Wheeler Aligner (v.0.7.12) with the default option (Li and Durbin, 2009), and alignment information was stored in Sequence Alignment Map (SAM) format. SAMtools (v.1.0) software was used to convert SAM files to BAM files before merging them for individual samples (Li et al., 2009). Picard-tools-1.117 (<https://broadinstitute.github.io/picard/>) MarkDuplicates was used to remove duplicate reads after mapping. The Genome Analysis Toolkit (v.3.3.0) was used for global realignment of reads around indels with the RealignerTargetCreator/IndelRealigner option (McKenna et al., 2010). Genomic variants were called for each accession by Genome Analysis Toolkit HaplotypeCaller (v.3.3.0). A set of raw variants was removed upon meeting the following filtering criteria: “DP < 200 || DP > 1000 || QD < 20.0 || FS > 10.0 || MQ < 40.0 || ReadPosRankSum < -8.0 || MQRankSum < -12.5” for SNPs and “DP < 200 || DP > 1000 || QD < 20.0 || FS > 20.0 || ReadPosRankSum < -20.0” for indels. Captured variants were annotated with iTools (Dinov et al., 2008).

SNP validation

We randomly selected 35 SNPs for PCR-based Sanger sequencing to examine the accuracy of the called SNPs (Supplemental Table 6). Amplifications were attempted in 12 accessions, including three from Turkey, two from Lebanon, two from Jordan, two from Syria, one from Iraq, one from Iran, and one from Armenia. Primers were designed flanking the focal SNP; at least one sample was predicted to be identical to the reference allele, and another sample was predicted to be identical to the alternate allele.

Phylogenetic reconstruction and population structure analyses

To build a phylogenetic tree, we converted the called SNPs into binary bed format and filtered them for MAF >0.05 and missing data <10% using PLINK (v.1.90) software (Purcell et al., 2007). A subset of 1 024 432 SNPs was used to calculate pairwise genetic distances among the 59 *T. urartu* accessions. A neighbor-joining tree was constructed using PHYLIP (v.3.695) (Cummings, 2004) according to the distance matrix. MEGA7 (Kumar et al., 2016) was used to visualize the phylogenetic trees. We also investigated the population structure using ADMIXTURE (v.1.3.0) based on the same dataset (Alexander et al., 2009). The program was run for several *K* values ranging from 2 to 7 with 10 000

iterations to determine the group membership of each accession. According to Evanno et al. (2010), $K = 2$ was chosen as the most likely group number with a higher ΔK calculated by STRUCTURE v.2.3.4 (Falush et al., 2003). In addition, we performed PCA based on the covariance matrix using EIGENSOFT (v.5.0.1) software (Price et al., 2006), and the first two eigenvectors were plotted in two dimensions. To illustrate the evolutionary history of the A subgenome, einkorn wheat from this study and the A subgenome of bread wheat (Chinese Spring) (Eversole, 2013) were used as outgroups for phylogenetic analysis. Clean reads from the three *Triticeae* species were aligned to the *T. urartu* reference genome using BWA and SAMtools, and 66 245 397 SNPs within the orthologous regions were extracted. A neighbor-joining tree was then generated based on the same method, in which the genotypes of einkorn wheat and bread wheat provided outgroup information at the corresponding positions.

LD analysis

To evaluate the LD level, the correlation coefficient (r^2) between SNPs was calculated using PopLDdecay v.3.30 (Zhang et al., 2018) with -MaxDist 500 -MAF 0.01 -Miss 0.1, and the half-maximum r^2 ($r^2 = 0.16$) decay distance was separately read as the LD length for the *T. urartu* populations. Simultaneously, LD blocks were defined by PLINK (v.1.90) software, and we divided the genome into small regions of 1 Mb in length. The parameters were set as follows: -geno 0.1 -maf 0.05 -blocks no-pheno-req -blocks-max-kb 500 -blocks-min-maf 0.05. The LD block definition was compared across the two groups for each individual chromosome.

TreeMix analysis

Admixture and gene flow among different geographic quadrants within the *T. urartu* gene pool were modeled in TreeMix v.1.13 (Figure 6A) (Pickrell and Pritchard, 2012). We divided the *T. urartu* accessions into seven regional groups according to their habitat population structure (Supplemental Figure 17; Supplemental Table 23). Segregating SNPs across the seven chromosomes were filtered with a genotyping rate >90% and an MAF >5%, from which 16.86 M were selected to construct an ML tree. For each TreeMix run, 500 SNP sites corresponding to a ~100-kb genomic region were analyzed as a block to account for possible LD effects. We considered admixture scenarios with $m = 0$ to $m = 5$ migration events, and the model fit for each migration was examined by estimating the proportion of variance between populations. If there were migration edges in the tree, then these would be colored according to their weight. To further investigate the origin of the A subgenome of polyploid wheat, we also built an admixture tree using the A subgenome of Chinese Spring as an outgroup. The bootstrap values on these trees were set to 1000 replicates.

Demographic analysis

The sequential Markov coalescent implemented in SMC++ (v.1.14.0) was used to infer the demographic history of the *T. urartu* accessions (Terhorst et al., 2017). Compared with the multiple sequentially Markovian coalescent and pairwise sequentially Markovian coalescent methods (Li and Durbin, 2011; Schiffels and Durbin, 2014), SMC++ could provide more accurate estimates, especially for recent population sizes. It allows multiple samples to be analyzed with genome-wide SNPs and does not require the genomes to be phased. To normalize the population size, we randomly selected ten different samples from the MT and EMC groups and then ran SMC++ software using default parameter settings. For each group, every sample was set to fit the demographic model, with all the genomic sequences used as input under the optimization algorithm of Powell. The split time between the MT and EMC groups was estimated with the command "SMC++ split". The analysis was performed assuming a mutation rate of 5.5×10^{-9} per site per generation and a constant generation time of 1 year (Dvorak and Akhunov, 2005; Chao et al., 2009).

Genetic diversity and F_{ST} calculation

The final high-quality variants were used to estimate nucleotide diversity (π) and Watterson's estimator (θ_W) for the whole *T. urartu* panel and for populations assigned by structure analysis and geographic quadrant. Among these, nucleotide diversity (π) was defined as the average number of differences per site between two randomly selected sequences (Tajima, 1983). Watterson's estimator (θ_W) was calculated using the equation $\theta_W = 4N_e\mu$, where N_e is the effective population size and μ is the per-generation mutation rate of the population (Watterson, 1975). The fixation index (F_{ST}) between the EMC and MT groups was calculated using Hudson's estimator (Bhatia et al., 2013). A sliding window of 500 kb with a 100-kb step size across the *T. urartu* genome was assigned for window calculation of π , θ_W , and F_{ST} , and only the windows comprising the ≥ 200 -kb effective covered region were retained for further analysis, according to in-house Perl scripts from a previous study (Zeng et al., 2018).

Scanning for genome-wide selective sweep signals

To identify genomic regions subjected to geographic adaptation, the nucleotide diversity ratio ($\ln \pi_{EMC}/\pi_{MT}$) was first analyzed with 500-kb windows and 100-kb steps across the genome, where π_{EMC} and π_{MT} are the π values of the EMC and MT populations, respectively. We then calculated the genome-wide distribution of F_{ST} values between the two populations using the same sliding-window approach. For convenience, the F_{ST} values were Z transformed, and the π ratios were ln transformed. Using a significance level of the top 5% for both $Z(F_{ST})$ and $\ln(\pi$ ratio) ($Z > 2.31$ and π ln ratio > 0.43 ; Figure 5B; Supplemental Table 14), a total of 70 potential selective sweep regions were identified (with an average size of 1.60 Mb, ranging from 0.10 to 8.30 Mb). They overlapped with 404 candidate genes and were used in the subsequent analysis and discussion. We also computed Tajima's D value (Tajima, 1989) and average r^2 in each window (RRSelection, <https://github.com/BGI-shenzhen/RRSelection>) for the two populations to test whether the candidate selective-sweep regions contained an excess of singleton polymorphisms. Specifically, this extended LD method calculated the sliding window mean r^2 for two groups across the genome and selected candidate selective regions with the greatest difference according to the top distribution of the measured mean r^2 (Z-test p value). The result confirmed the occurrence of selective sweeps in the identified regions from another perspective.

Two cross-population comparison statistical methods, XP-CLR and XP-EHH, were also used to scan the genome (Sabeti et al., 2007; Chen et al., 2010). Selection signatures under geographic adaptation were evaluated at the whole-genome level between the two populations. The composite likelihood approach (XP-CLR) was performed using a 0.005-cM sliding window with a 200-kb grid size, which required a genetic map of *T. urartu* (Ling et al., 2018). All individual SNPs were assigned to a position by assuming uniform recombination between mapped markers. We fixed 200 SNPs assayed in each window to ensure comparability of the composite likelihood and downweighted pairs of SNPs in high LD ($r^2 > 0.90$) to minimize the effect of dependence. Subsequently, the XP-CLR score was calculated, with the top 5% cutoff threshold of the genome considered to represent adaptation-selected regions. The XP-EHH statistic was used to confirm the selective sweeps by assessing extended haplotype homozygosity between the two populations.

Gene functional annotation and gene set analysis

All candidate selective genes were annotated using the Blast2GO program (Conesa and Götz, 2008). GO and functional classification analyses of these genes were performed using AmiGO (The Gene Ontology Consortium, 2015). We also uploaded the candidate genes to the KEGG Automated Annotation Server (KAAS, <http://www.genome.jp/tools/kaas/>) to obtain the KEGG pathways. The χ^2 method was then used to test the statistical significance of the enrichments. The p values were adjusted by FDR, and the cutoff was set to 0.05. The CDS of each

Plant Communications

selected gene was compared against the NCBI nr database to identify homologous genes in *A. thaliana*, rice, and maize.

Phenotyping

For phenotyping, the 221 accessions were planted in two consecutive years (2013–2014 and 2014–2015 cropping seasons) at three agricultural experiment stations: the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing; Henan Agricultural University, Zhengzhou; and Dezhou Academy of Agricultural Sciences, Dezhou. Each accession was planted in a single 2-m row with 40-cm spacing between rows. The heading date for each accession was recorded as the days from sowing to the time when half of the spikes emerged from the flag leaf. Plant height was measured during the stage when the maximum height was achieved (from ground level to the apex of the spike, excluding awns). Spike length and spikelet number per spike were determined by measuring three spikes on each individual plant and counting individual spikelets per spike. The grain length, grain width, and grain length/width ratio of each accession were scanned after harvest using an electronic digital caliper. Thousand-grain weight was obtained from the average of two 1000-kernel weights (Supplemental Tables 26–28).

GWAS

Combining the WGS and GBS data, we obtained a final set of 455 644 SNPs. After filtering, 149 030 high-quality SNPs (MAF > 0.05 and missing data < 40%) were used to perform a GWAS on eight agronomic and grain traits in 221 accessions. The association analyses were performed with TASSEL 5.0 software (Bradbury et al., 2007) via a compressed MLM using the Q matrix and K matrix. The Q matrix was calculated from structure analysis for the stratification control, whereas the K matrix was calculated with the TASSEL plugin “-KinshipPlugin -method Centered_IBS”. The parameters for GWAS were “run_pipeline.pl -Xmx10G -fork1 -h \$hmp -filterAlign -filterAlignMin-Freq 0.05 -fork2 -r \$phenotype -fork3 -q \$Q_matrix -excludeLastTrait -fork4 -k \$K_matrix -combine5 -input1 -input2 -input3 -intersect -combine6 -input5 -input4 -mlm -mlmVarCompEst P3D -mlmCompressionLevel Optimum -export \$output -runfork1 -runfork2 -runfork3 -runfork4”. Manhattan and quantile-quantile plots were then drawn in R using the qqman package (Turner, 2014). The genome-wide significance threshold was determined using permutation-based, FDR-adjusted p values, for which the permutation tests were repeated 1000 times. To detect the candidate gene in the significantly associated region (chromosome Tu2: 2,752,902–2,691,906), we aligned paired-end reads containing causative variants for the high and short accessions that were resequenced against the reference genome using BWA and SAMtools. We measured the LD level around the significant SNPs by LD block analysis using Haploview software (Barrett et al., 2005). Candidate genes were subsequently used for a BLAST query in NCBI to obtain the coordinates and percentage similarity of orthologous genes.

Expression analysis

The candidate genes in *T. urartu* accessions were analyzed by quantitative real-time PCR. Specifically, at the heading date, leaves for *HD16* and *FT1* analyses and young panicles for *SD37* analysis were collected for RNA extraction. Total RNA was extracted from the liquid nitrogen-frozen samples using TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. First-strand cDNA was synthesized with Hifair III 1st Strand cDNA Synthesis SuperMix (Yeasen Biotech, Shanghai, China). Gene-specific primers were designed using Primer Premier 5.0 software (<http://www.premierbiosoft.com/primerdesign/>). Quantitative real-time PCR was performed with Hieff qPCR SYBR Green Master Mix (Yeasen Biotech) using the QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The primers used for the *HD16* ortholog (TuG1812G0200002519) were 5'-GGCTACCAGGAGATACCA-3' and 5'-ATTCTGATGGCCTTGACGC-3'; those for the *FT1* ortholog (TuG1812G0300001662) were 5'-TCGTACGGATGCATGTCCAG-3' and

Whole-genome resequencing of *Triticum urartu*

5'-TGGCAGTTGAAGTAGACGGC-3'; and those for the *SD37* ortholog (TuG1812G0200000045) were 5'-TGACCTGCAACCTCGTCTTC-3' and 5'-GAACCGGTGATGGTCTAC-3'. The expression of *TuActin* (5'-CAGGCCGTTCTGCTTGTGTA-3' and 5'-CAGGCCGTTCTGCTTGTGTA-3') was used as an internal control. The relative expression of target genes was calculated based on the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Three independent replicates were performed for each sample.

ACCESSION NUMBERS

All the sequencing data generated in this study have been deposited in the public database of the China National GeneBank (<https://db.cngb.org/cnsa/>) under accession numbers CNP0000699 for WGS of *Triticum urartu* and einkorn wheats (<https://db.cngb.org/search/?q=CNP0000699>) and CNP0000775 for GBS data of *Triticum urartu* (<https://db.cngb.org/search/?q=CNP0000775>).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This research was financially supported by the National Natural Science Foundation of China (31871617) and the Ministry of Science and Technology of the People's Republic of China (2016YFD0102002 and 2011AA100104).

AUTHOR CONTRIBUTIONS

A.Z., D.L., and S.Z. conceived and designed the research. X.W., Y.L., W.Y., J.S., and X.L. carried out the field work and collected samples. X.W., Y.H., W.H., C.Z., F.Z., and S.Z. carried out sequencing and data analysis. X.W., K.Y., and D.L. prepared the manuscript. S.Z., L.K., and H.L. revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank the National Small Grain Collection (USDA-ARS) and the International Center for Agricultural Research in the Dry Areas (ICARDA) for providing the seed samples. The authors declare that they have no competing interests.

Received: January 2, 2022

Revised: April 23, 2022

Accepted: May 30, 2022

Published: June 1, 2022

REFERENCES

- Abbo, S., and Gopher, A. (2017). Near eastern plant domestication: a history of thought. *Trends Plant Sci.* **22**:491–511. <https://doi.org/10.1016/j.tplants.2017.03.010>.
- Adamski, N.M., Borrill, P., Brinton, J., Harrington, S.A., Marchal, C., Bentley, A.R., Bovill, W.D., Cattivelli, L., Cockram, J., Contreras-Moreira, B., et al. (2020). A roadmap for gene functional characterisation in crops with large genomes: lessons from polyploid wheat. *Elife* **9**:e55646. <https://doi.org/10.7554/elife.55646>.
- Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**:1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Alvarez, J.B., Caballero, L., Nadal, S., Ramirez, M., and Martin, A. (2009). Development and gluten strength evaluation of introgression lines of *Triticum urartu* in durum wheat. *Cereal Res. Commun.* **37**:243–248. <https://doi.org/10.1556/crc.37.2009.2.11>.
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**:93–97. <https://doi.org/10.1126/science.aan0032>.

- El Baidouri, M., Murat, F., Veyssièrè, M., Molinier, M., Flores, R., Burlot, L., Alaux, M., Quesneville, H., Pont, C., and Salse, J. (2017). Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytol.* **213**:1477–1486. <https://doi.org/10.1111/nph.14113>.
- Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**:263–265. <https://doi.org/10.1093/bioinformatics/bth457>.
- Bhatia, G., Patterson, N., Sankararaman, S., and Price, A.L. (2013). Estimating and interpreting F_{ST} : the impact of rare variants. *Genome Res.* **23**:1514–1521. <https://doi.org/10.1101/gr.154831.113>.
- Bhuiyan, N.H., Liu, W.P., Liu, G.S., Selvaraj, G., Wei, Y.D., and King, J. (2007). Transcriptional regulation of genes involved in the pathways of biosynthesis and supply of methyl units in response to powdery mildew attack and abiotic stresses in wheat. *Plant Mol. Biol.* **64**:305–318. <https://doi.org/10.1007/s11103-007-9155-x>.
- Bradbury, P.J., Zhang, Z.W., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>.
- Brunazzi, A., Scaglione, D., Talini, R.F., Miculan, M., Magni, F., Poland, J., Enrico Pè, M., Brandolini, A., and Dellacqua, M. (2018). Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the Fertile Crescent. *Plant J.* **94**:670–684. <https://doi.org/10.1111/tpj.13888>.
- Cao, H., Li, X.Y., Wang, Z., Ding, M., Sun, Y.Z., Dong, F.Q., Chen, F.Y., Liu, L., Doughty, J., Li, Y., et al. (2015). Histone H2B Monoubiquitination mediated by HISTONE MONOUBIQUITINATION1 and HISTONE MONOUBIQUITINATION2 is involved in anther development by regulating tapetum degradation-related genes in rice. *Plant Physiol* **168**:1389–1405. <https://doi.org/10.1104/pp.114.256578>.
- Chao, S., Zhang, W.J., Akhunov, E., Sherman, J.D., Ma, Y.Q., Luo, M.C., and Dubcovsky, J. (2009). Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. *Mol. Breeding* **23**:23–33. <https://doi.org/10.1007/s11032-008-9210-6>.
- Charmet, G. (2011). Wheat domestication: lessons for the future. *C. R. Biol* **334**:212–220. <https://doi.org/10.1016/j.crv.2010.12.013>.
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* **20**:393–402. <https://doi.org/10.1101/gr.100545.109>.
- Chen, L.J., Diao, Z.Y., Specht, C., and Sung, Z.R. (2009). Molecular evolution of VEF-domain-containing PcG genes in plants. *Mol. Plant* **2**:738–754. <https://doi.org/10.1093/mp/ssp032>.
- Chen, Y., Shi, C.M., Huang, Z.Y., Zhang, Y., Li, S., Li, Y.R., Ye, J., Yu, C., Li, Z., Zhang, X., et al. (2018). SOAPnuke: a MapReduce acceleration supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**:1–6. <https://doi.org/10.1093/gigascience/gix120>.
- Cheng, H., Liu, J., Wen, J., Nie, X.J., Xu, L.H., Chen, N.B., Li, Z.X., Wang, Q.L., Zheng, Z.Q., Li, M., et al. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biol.* **20**:1–16. <https://doi.org/10.1186/s13059-019-1744-x>.
- Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**:619832. <https://doi.org/10.1155/2008/619832>.
- Cummings, M.P. (2004). PHYLIP (Phylogeny Inference Package) (John Wiley & Sons, Inc.), pp. 164–166.
- Dahleen, L.S., Tyagi, N., Brown, R.H., Morgan, W.C., and Morgan, W.C. (2012). Developing tools for investigating the multiple roles of ethylene: identification and mapping genes for ethylene biosynthesis and reception in barley. *Mol. Genet. Genomics* **287**:793–802. <https://doi.org/10.1007/s00438-012-0716-6>.
- Dempewolf, H., Baute, G.J., Anderson, J.E., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* **57**:1070–1082. <https://doi.org/10.2135/cropsci2016.10.0885>.
- Dickinson, H., Costa, L., and Gutierrez-Marcos, J. (2012). Epigenetic neofunctionalisation and regulatory gene evolution in grasses. *Trends Plant Sci.* **17**:389–394. <https://doi.org/10.1016/j.tplants.2012.04.002>.
- Dinov, I.D., Rubin, D.L., Lorensen, W.E., Dugan, J.M., Ma, J., Murphy, S.N., Kirschner, B., Bug, W.J., Sherman, M.A., Floratos, A., et al. (2008). iTools: a framework for classification, categorization and integration of computational biology resources. *PLoS One* **3**:e2265. <https://doi.org/10.1371/journal.pone.0002265>.
- Dvorak, J., and Akhunov, E.D. (2005). Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. *Genetics* **171**:323–332. <https://doi.org/10.1534/genetics.105.041632>.
- Dvořák, J., Terlizzi, P.d., Zhang, H.B., and Resta, P. (1993). The evolution of polyploid wheats: identification of the A genome donor species. *Genome* **36**:21–31. <https://doi.org/10.1139/g93-004>.
- Elschire, R.J., Glaubitz, J.C., Sun, Q.Y., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**:e19379. <https://doi.org/10.1371/journal.pone.0019379>.
- Evanno, G., Regnaut, S., and Goudet, J. (2010). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**:2611–2620. <https://doi.org/10.1111/j.1365-294x.2005.02553.x>.
- Evans, L.M., Slavov, G.T., Rodgersmelnick, E., Martin, J., Ranjan, P., Muchero, W., Brunner, A.M., Schackwitz, W., Gunter, L.E., Chen, J.G., et al. (2014). Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat. Genet.* **46**:1089–1096. <https://doi.org/10.1038/ng.3075>.
- Eversole, K. (2013). The International wheat genome sequencing consortium (IWGSC). *Plant Animal Genome*. <http://www.wheatgenome.org>.
- Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**:1567–1587. <https://doi.org/10.1093/genetics/164.4.1567>.
- Fang, Z.M., Xia, K.F., Yang, X., Grotemeyer, M.S., Meier, S., Rentsch, D., Xu, X.L., and Zhang, M.Y. (2013). Altered expression of the *PTR/NRT1* homologue *OsPTR9* affects nitrogen utilization efficiency, growth and grain yield in rice. *Plant Biotech. J.* **11**:446–458. <https://doi.org/10.1111/pbi.12031>.
- Gustafson, P., Raskina, O., Ma, X., and Nevo, E. (2009). Wheat evolution, domestication, and improvement. In *Wheat: Science and Trade*, B.F. Carver, ed. (Wiley Blackwell), pp. 3–30.
- Haldorsen, S., Akan, H., Celik, B., and Heun, M. (2011). The climate of the Younger Dryas as a boundary for Einkorn domestication. *Veg. Hist. Archaeobot.* **20**:305–318. <https://doi.org/10.1007/s00334-011-0291-5>.
- Hammer, K., Filatenko, A.A., and Korzun, V. (2000). Microsatellite markers—a new tool for distinguishing diploid wheat species. *Genet. Resour. Crop Ev.* **47**:497–505. <https://doi.org/10.1023/a:1008761928959>.
- Hong, Z., Ueguchi-Tanaka, M., Shimizu-Sato, S., Inukai, Y., Fujioka, S., Shimada, Y., Takatsuto, S., Agetsuma, M., Yoshida, S., Watanabe, Y., et al. (2002). Loss-of-function of a rice brassinosteroid biosynthetic enzyme, C-6 oxidase, prevents

- the organized arrangement and polar elongation of cells in the leaves and stem. *Plant J.* **32**:495–508. <https://doi.org/10.1046/j.1365-3113x.2002.01438.x>.
- Hori, K., Ogiso-Tanaka, E., Matsubara, K., Yamanouchi, U., Ebana, K., and Yano, M. (2013). *Hd16*, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant J.* **76**:36–46. <https://doi.org/10.1111/tpj.12268>.
- Huang, L., Raats, D., Sela, H., Klymiuk, V., Lidzbarsky, G., Feng, L., Krugman, T., and Fahima, T. (2016). Evolution and adaptation of wild emmer wheat populations to biotic and abiotic stresses. *Annu. Rev. Phytopathol.* **54**:279–301. <https://doi.org/10.1146/annurev-phyto-080614-120254>.
- Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppeler, S.M., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **44**:808–811. <https://doi.org/10.1038/ng.2309>.
- International Wheat Genome Sequencing Consortium. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**:eaar7191. <https://doi.org/10.1126/science.aar7191>.
- Jia, J.Z., Zhao, S.C., Kong, X.Y., Li, Y.R., Zhao, G.Y., He, W.M., Appels, R., Pfeifer, M., Tao, Y., Zhang, X.Y., et al. (2013). *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**:91–95. <https://doi.org/10.1038/nature12028>.
- Jones, M.K., Allaby, R.G., Brown, T.A., Hole, F., Heun, M., Borghi, B., and Salamini, F. (1998). Wheat domestication. *Science* **279**:302. <https://doi.org/10.1126/science.279.5349.302c>.
- Kasim, W.A., Osman, M.E., Omar, M.N., Abd El-Daim, I.A., Bejai, S., and Meijer, J. (2013). Control of drought stress in wheat using plant-growth-promoting bacteria. *J. Plant Growth Regul.* **32**:122–130. <https://doi.org/10.1007/s00344-012-9283-7>.
- Keller, M., Karutz, C., Schmid, J.E., Stamp, P., Winzeler, M., Keller, B., and Messmer, M.M. (1999). Quantitative trait loci for lodging resistance in a segregating wheat × spelt population. *Theor. Appl. Genet.* **98**:1171–1182. <https://doi.org/10.1007/s001220051182>.
- Klinge, J.A., and Fall, P. (2010). Archaeobotanical inference of Bronze Age land use and land cover in the eastern Mediterranean. *J. Arch. Sci.* **37**:2622–2629. <https://doi.org/10.1016/j.jas.2010.05.022>.
- Knapp, A.B. (1992). Bronze Age Mediterranean island cultures and the ancient near east. *Biblic. Archaeol.* **55**:52–72. <https://doi.org/10.2307/3210346>.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
- Kwon, C.T., Kim, S.H., Kim, D., and Paek, N.C. (2015). The rice floral repressor *Early flowering1* affects spikelet fertility by modulating gibberellin signaling. *Rice* **8**:23. <https://doi.org/10.1186/s12284-015-0058-1>.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* **475**:493–496. <https://doi.org/10.1038/nature10231>.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.T., Abecasis, G.R., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Lian, W.W., Tang, Y.M., Gao, S.Q., Zhang, Z., Zhao, X., and Zhao, C.P. (2012). Phylogenetic analysis and expression patterns of the MAPK gene family in wheat (*Triticum aestivum* L.). *J. Integr. Agr.* **11**:1227–1235. [https://doi.org/10.1016/s2095-3119\(12\)60119-1](https://doi.org/10.1016/s2095-3119(12)60119-1).
- Ling, H.Q., Ma, B., Shi, X.L., Liu, H., Dong, L.L., Sun, H., Cao, Y.H., Gao, Q., Zheng, S.S., Li, Y., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* **557**:424–428. <https://doi.org/10.1038/s41586-018-0108-0>.
- Ling, H.Q., Zhao, S.C., Liu, D.C., Wang, J.Y., Sun, H., Zhang, C., Fan, H.J., Li, D., Dong, L.L., Tao, Y., et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**:87–90. <https://doi.org/10.1038/nature11997>.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods* **25**:402–408. <https://doi.org/10.1006/meth.2001.1262>.
- Luo, G.B., Zhang, X.F., Zhang, Y.L., Yang, W.L., Li, Y.W., Sun, J.Z., Zhan, K.H., Zhang, A.M., and Liu, D.C. (2015). Composition, variation, expression and evolution of low-molecular-weight glutenin subunit genes in *Triticum urartu*. *BMC Plant Biol.* **15**:322. <https://doi.org/10.1186/s12870-014-0322-3>.
- Luo, G.B., Shen, L.S., Song, Y.H., Yu, K., Ji, J.J., Zhang, C., Yang, W.L., Li, X., Sun, J.Z., Zhan, K.H., et al. (2021). The MYB family transcription factor *TuODORANT1* from *Triticum urartu* and the homolog *TaODORANT1* from *Triticum aestivum* inhibit seed storage protein synthesis in wheat. *Plant Biotechnol. J.* **19**:1863–1877. <https://doi.org/10.1111/pbi.13604>.
- Luo, H., Song, F., and Zheng, Z. (2005). Overexpression in transgenic tobacco reveals different roles for the rice homeodomain gene *OsBIHD1* in biotic and abiotic stress responses. *J. Exp. Bot.* **56**:2673–2682. <https://doi.org/10.1093/jxb/eri260>.
- Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T.T., Wang, L., Wang, Y., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **551**:498–502. <https://doi.org/10.1038/nature24486>.
- Maccacferri, M., Sanguineti, M.C., Corneti, S., Ortega, J.L.A., Salem, M.B., Bort, J., Deambrogio, E., Del Moral, L.F.G., Demontis, A., Elahmed, A., et al. (2008). Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics* **178**:489–511. <https://doi.org/10.1534/genetics.107.077297>.
- Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B.B.H., Steuernagel, B., Mayer, K.F.X., Olsen, O.A., et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**:1250092. <https://doi.org/10.1126/science.1250092>.
- McGovern, T.H. (2010). *The archaeology of global change: The impact of humans on their environment*, edited by C.L. Redman, S.R. James, P.R. Fish, and J. Daniel Rogers. *Am. Anthropol.* **110**:76–78.
- Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A.M., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.J., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**:1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Meyer, R.S., Choi, J.Y., Sanches, M., Plessis, A., Flowers, J.M., Amas, J., Dorph, K., Barretto, A., Gross, B.L., Fuller, D.Q., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* **48**:1083–1088. <https://doi.org/10.1038/ng.3633>.
- Moles, A.T., Warton, D.I., Warman, L., Swenson, N.G., Laffan, S.W., Zanne, A.E., Pitman, A.J., Hemmings, F.A., and Leishman, M.R. (2009). Global patterns in plant height. *J. Ecol.* **97**:923–932. <https://doi.org/10.1111/j.1365-2745.2009.01526.x>.

- Nallamilli, B.R.R., Zhang, J., Mujahid, H., Malone, B.M., Bridges, S.M., and Peng, Z.H. (2013). Polycomb group gene *OsFIE2* regulates rice (*Oryza sativa*) seed development and grain filling via a mechanism distinct from *Arabidopsis*. *PLoS Genet.* **9**:e1003322. <https://doi.org/10.1371/journal.pgen.1003322>.
- Pankin, A., Altmüller, J., Becker, C., and von Korff, M. (2018). Targeted resequencing reveals genomic signatures of barley domestication. *New Phytol.* **218**:1247–1259. <https://doi.org/10.1111/nph.15077>.
- Peng, J.H., Sun, D., and Nevo, E. (2011). Domestication evolution, genetics and genomics in wheat. *Mol. Breeding* **28**:281–301. <https://doi.org/10.1007/s11032-011-9608-4>.
- Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**:e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**:904–909. <https://doi.org/10.1038/ng1847>.
- Purcell, S., Neale, B.M., 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>.
- Qiu, Y.C., Zhou, R.H., Kong, X.Y., Zhang, S.S., and Jia, J.Z. (2005). Microsatellite mapping of a *Triticum urartu* Tum. derived powdery mildew resistance gene transferred to common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **111**:1524–1531. <https://doi.org/10.1007/s00122-005-0081-5>.
- Rasheed, A., and Xia, X.C. (2019). From markers to genome-based breeding in wheat. *Theor. Appl. Genet.* **132**:767–784. <https://doi.org/10.1007/s00122-019-03286-4>.
- Reif, J.C., Zhang, P., Dreisigacker, S., Warburton, M.L., van Ginkel, M., Hoisington, D.A., Bohn, M.O., and Melchinger, A.E. (2005). Wheat genetic diversity trends during domestication and breeding. *Theor. Appl. Genet.* **110**:859–864. <https://doi.org/10.1007/s00122-004-1881-8>.
- Ramamoorthy, R., Jiang, S.Y., and Ramachandran, S. (2011). *Oryza sativa* cytochrome P450 family member *OsCYP96B4* reduces plant height in a transcript dosage dependent manner. *PLoS One* **6**:e28069. <https://doi.org/10.1371/journal.pone.0028069>.
- Renssen, H., Mairesse, A., Goosse, H., Mathiot, P., Heiri, O., Roche, D.M., Nisancioglu, K.H., and Valdes, P.J. (2015). Multiple causes of the Younger Dryas cold period. *Nat. Geosci.* **8**:946–949. <https://doi.org/10.1038/ngeo2557>.
- Romero Navarro, J.A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., Preciado, E., Terron, A., Delgado, H.V., Vidal, V., et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**:476–480. <https://doi.org/10.1038/ng.3784>.
- Rouse, M.N., and Jin, Y. (2011). Stem rust resistance in A-genome diploid relatives of wheat. *Plant Dis.* **95**:941–944. <https://doi.org/10.1094/pdis-04-10-0260>.
- Roy, S.J., Tucker, E.J., and Tester, M. (2011). Genetic analysis of abiotic stress tolerance in crops. *Curr. Opin. Plant Biol.* **14**:232–239. <https://doi.org/10.1016/j.pbi.2011.03.002>.
- Russell, J., Mascher, M., Dawson, I.K., Kyriakidis, S., Calixto, C.P.G., Freund, F., Bayer, M., Milne, I., Marshall-Griffiths, T., Heinen, S., et al. (2016). Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat. Genet.* **48**:1024–1030. <https://doi.org/10.1038/ng.3612>.
- Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918. <https://doi.org/10.1038/nature06250>.
- Saghai-Marooif, M.A., Soliman, K.M., Jorgensen, R.A., and Allard, R.W.L. (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U S A* **81**:8014–8018. <https://doi.org/10.1073/pnas.81.24.8014>.
- Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**:919–925. <https://doi.org/10.1038/ng.3015>.
- Shen, L.S., Luo, G.B., Song, Y.H., Xu, J.Y., Ji, J.J., Zhang, C., Gregová, E., Gregová, E., Yang, W.L., Li, X., et al. (2020). A novel NAC family transcription factor *SPR* suppresses seed storage protein synthesis in wheat. *Plant Biotechnol. J.* **19**:992–1007. <https://doi.org/10.1111/pbi.13524>.
- Song, S.Y., Chen, Y., Liu, L., Wang, Y.W., Bao, S.J., Zhou, X., Teo, Z.W.N., Mao, C.Z., Gan, Y.B., and Yu, H. (2017). *OsFTIP1*-mediated regulation of florigen transport in rice is negatively regulated by the ubiquitin-like domain kinase *OsUbdKγ4*. *Plant Cell* **29**:491–507. <https://doi.org/10.1105/tpc.16.00728>.
- Song, Y.H., Luo, G.B., Shen, L.S., Yu, K., Yang, W.L., Li, X., Sun, J.Z., Zhan, K.H., Cui, D.Q., Liu, D.C., et al. (2020). *TubZIP28*, a novel bZIP family transcription factor from *Triticum urartu*, and *TabZIP28*, its homologue from *Triticum aestivum*, enhance starch synthesis in wheat. *New Phytol.* **226**:1384–1398. <https://doi.org/10.1111/nph.16435>.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**:437–460. <https://doi.org/10.1093/genetics/105.2.437>.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Tamiru, M., Undan, J.R., Takagi, H., Abe, A., Yoshida, K., Undan, J.Q., Natsume, S., Uemura, A., Saitoh, H., Matsumura, H., et al. (2015). A cytochrome P450, *OsDSS1*, is involved in growth and drought stress responses in rice (*Oryza sativa* L.). *Plant Mol. Biol.* **88**:85–99. <https://doi.org/10.1007/s11103-015-0310-5>.
- Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**:303–309. <https://doi.org/10.1038/ng.3748>.
- The Gene Ontology Consortium. (2015). Gene ontology consortium: going forward. *Nucleic Acids Res.* **43**:1049–1056.
- Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at bioRxiv. <https://doi.org/10.1101/005165>.
- Vatsiou, A.I., Bazin, E., and Gaggiotti, O.E. (2016). Detection of selective sweeps in structured populations: a comparison of recent methods. *Mol. Ecol.* **25**:89–103. <https://doi.org/10.1111/mec.13360>.
- Wang, G.F., Fan, R., Wang, X., Wang, D., and Zhang, X. (2015). *TaRAR1* and *TaSGT1* associate with *TaHsp90* to function in bread wheat (*Triticum aestivum* L.) seedling growth and stripe rust resistance. *Plant Mol. Biol.* **87**:577–589. <https://doi.org/10.1007/s11103-015-0298-x>.
- Wang, H., and Wang, H. (2015). Phytochrome signaling: time to tighten up the loose ends. *Mol. Plant* **8**:540–551. <https://doi.org/10.1016/j.molp.2014.11.021>.
- Wang, X., Luo, G.B., Yang, W.L., Li, Y.W., Sun, J.Z., Zhan, K.H., Liu, D.C., and Zhang, A.M. (2017). Genetic diversity, population structure and marker-trait associations for agronomic and grain traits

Plant Communications

- in wild diploid wheat *Triticum urartu*. *BMC Plant Biol.* **17**:112. <https://doi.org/10.1186/s12870-017-1058-7>.
- Watterson, G.A.** (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**:256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9).
- Wright, H.E.** (1976). The environmental setting for plant domestication in the near east. *Science* **194**:385–389. <https://doi.org/10.1126/science.194.4263.385>.
- Xu, P., Xiang, Y., Zhu, H.L., Xu, H.B., Zhang, Z.Z., Zhang, C.Q., Zhang, L.X., and Ma, Z.Q.** (2009). Wheat cryptochromes: subcellular localization and involvement in photomorphogenesis and osmotic stress responses. *Plant Physiol* **149**:760–774. <https://doi.org/10.1104/pp.108.132217>.
- Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F.Y., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., et al.** (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotech.* **30**:105–111. <https://doi.org/10.1038/nbt.2050>.
- Zeder, M.A.** (2011). The origins of agriculture in the near east. *Curr. Anthropol.* **52**:S221–S235. <https://doi.org/10.1086/659307>.
- Zeng, X.Q., Guo, Y., Xu, Q.J., Mascher, M., Guo, G.G., Li, S.C., Mao, L.K., Liu, Q.F., Xia, Z.F., Zhou, J.H., et al.** (2018). Origin and evolution of qingke barley in Tibet. *Nat. Commun.* **9**:5433. <https://doi.org/10.1038/s41467-018-07920-5>.
- Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L.** (2018). Poplddecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**:1786–1788. <https://doi.org/10.1093/bioinformatics/bty875>.
- Zhang, J., Liu, X.Q., Li, S.Y., Cheng, Z.K., and Li, C.Y.** (2014). The rice semi-dwarf mutant *sd37*, caused by a mutation in *CYP96B4*, plays an important role in the fine-tuning of plant growth. *PLoS One* **9**:e88068. <https://doi.org/10.1371/journal.pone.0088068>.
- Zhang, Y.L., Luo, G.B., Liu, D.C., Wang, D.Z., Yang, W.L., Sun, J.Z., Zhang, A.M., and Zhan, K.H.** (2015). Genome-transcriptome- and proteome-wide analyses of the gliadin gene families in *Triticum urartu*. *PLoS One* **10**:e0131559. <https://doi.org/10.1371/journal.pone.0131559>.
- Zhang, Z.B., Zhang, J.W., Chen, Y.J., Li, R.F., Wang, H.Z., and Wei, J.H.** (2012). Genome-wide analysis and identification of HAK potassium transporter gene family in maize (*Zea mays* L.). *Mol. Biol. Rep.* **39**:8465–8473. <https://doi.org/10.1007/s11033-012-1700-2>.
- Zhao, L., Tan, L.B., Zhu, Z.F., Xiao, L.T., Xie, D.X., and Sun, C.Q.** (2015). *<scp>PAY</scp> 1* improves plant architecture and enhances grain yield in rice. *Plant J.* **83**:528–536. <https://doi.org/10.1111/tpj.12905>.
- Zhou, Z.K., Jiang, Y., Wang, Z., Gou, Z.H., Lyu, J., Li, W.Y., Yu, Y.J., Shu, L.P., Zhao, Y.J., Ma, Y.M., et al.** (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotech.* **33**:408–414. <https://doi.org/10.1038/nbt.3096>.
- Zhou, Y., Zhao, X.B., Li, Y.W., Xu, J., Bi, A.Y., Kang, L.P., Xu, D.X., Chen, H.F., Wang, Y., Wang, Y.G., et al.** (2020). *Triticum* population sequencing provides insights into wheat adaptation. *Nat. Genet.* **52**:1412–1422. <https://doi.org/10.1038/s41588-020-00722-w>.
- Zou, X.H., Qin, Z.R., Zhang, C.Y., Liu, B., Liu, J., Zhang, C.S., Lin, C.T., Li, H.Y., and Zhao, T.** (2015). Over-expression of an S-domain receptor-like kinase extracellular domain improves panicle architecture and grain yield in rice. *J. Exp. Bot.* **66**:7197–7209. <https://doi.org/10.1093/jxb/erv417>.