Title:

Population genomics unravels the Holocene history of *Triticum-Aegilops* species

Authors

Xuebo Zhao[1,2,#], Yafei Guo[1,2,#], Lipeng Kang[1,2], Aoyue Bi[1,2], Daxing Xu[1,2], Zhiliang Zhang[1,2], Jijin Zhang[1,2], Xiaohan Yang[1,2], Jun Xu[1,2], Song Xu[1,2], Xinyue Song[1,3], Ming Zhang[1,2], Yiwen Li[1], Philip Kear[4], Jing Wang[1], Changbin Yin[1], Zhiyong Liu[1,2], Xiangdong Fu[1,2], Fei Lu[1,2,5,*]

Affiliations

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China.

[2]University of Chinese Academy of Sciences, Beijing, China.

[3]State Key Laboratory of Conservation and Utilization of Bio-Resources in Yunnan and Center for Life Sciences, School of Life Sciences, Yunnan University, Kunming, China

[4]International Potato Center-China

[5]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.

[#]These authors contributed equally to this work.

[*]Corresponding author: flu@genetics.ac.cn (F.L.)

Abstract

Deep knowledge of crop biodiversity is essential to improve global food security. Despite bread wheat serving as a keystone crop worldwide, the population history of bread wheat and its wild relatives (a.k.a. wheats) remains elusive. By analyzing whole-genome sequences of 795 wheats, we found that bread wheat originated southwest of the Caspian Sea ~11,700 years ago and underwent a slow speciation process, lasting ~3,300 years due to persistent gene flow from wild relatives. Soon after, bread wheat spread across Eurasia and reached Europe, South Asia, and East Asia ~7,000 to ~5,000 years ago, shaping a diversified but occasionally convergent adaptive landscape of bread wheat in novel environments. Opposite to cultivated wheat, wild wheat populations have declined by ~82% in the past ~2,000 years due to the food choice shift of humans, and likely continue to drop because of the changing climate. These findings will guide future efforts in protecting and utilizing wheat biodiversity to improve global food security.

Introduction

Climate change and the growing population are putting global food security at risk—the world crop production is projected to be inadequate by 2050[1]. While various adaptive strategies[2] and technologies[3] of plant breeding have been proposed to address the challenge, many of these opportunities lie in crop biodiversity, which preserves tremendous pre-adapted and beneficial alleles to develop productive, nutritious, stress-resilient, and sustainable crop varieties[4]. An in-depth understanding of cultivated crops and their wild relatives is central to integrating genetic resources and breeding methods effectively.

Bread wheat (*Triticum aestivum* ssp. *aestivum*, $2n = 6x = 42$, AABBDD) is one of the world's most important crops, providing ~20% calories and protein for the human diet[5]. Meanwhile, bread wheat and its relatives, such as domesticated einkorn (*T. monococcum* ssp. *monococcum*, AA) and domesticated emmer (*T. turgidum* ssp. *dicoccum*, AABB), were among the first crops bringing forth agriculture and subsequent civilization[6]. Due to the economic and cultural importance of these ancient crops, the evolutionary history of *Triticum* and *Aegilops* species, the two clades giving rise to modern bread wheat through polyploidization[7], has been of great interest to both scientists[7–9] and the public[10,11]. Fueled by the landmark bread wheat reference genome[12], recent studies have reconstructed the phylogeny of *Triticum-Aegilops* species[13,14], characterized the population structure of modern wheat[9,13,15–17], and identified historical gene flow from wild populations to bread wheat[13,14,16,18,19]. However, the population history of wheats (bread wheat and its wild relatives, or *Triticum-Aegilops* species) is largely incomplete, particularly the spatiotemporal dynamics of bread wheat emergence and dispersal, together with the genetic and ecological interaction between bread wheat and its wild relatives, remain elusive[6,20,21].

Here we performed a genus-level sampling of *Triticum-Aegilops* species and conducted whole-genome sequence analyses to disentangle the deep past of wheats since the rise of agriculture ~10,000 years ago. The paralleled reconstruction of demographic histories of both cultivated and wild wheats provided the first example of the Holocene evolution of the entire gene pool appertaining to a crop species, insights from which will benefit biodiversity conservation and breeding of many crops.

Results

Genomic data of *Triticum-Aegilops* populations

We collected whole-genome sequencing data of 795 accessions, including 745 accessions from publicly available data set[13,17,18], and 50 newly sequenced accessions in this study to complete the sampling of wild relatives of bread wheat. These highly diverse accessions are from 6 species and 25 subspecies in the genera *Triticum* and *Aegilops* (Fig. 1), representing a wide range of geographic distribution (73 countries, Supplementary Fig. 1), comprehensive ploidy levels (diploid, tetraploid, and hexaploid) and genome types (AA, BB/SS, AABB, AABBDD, and DD) related to the A, B, and D subgenomes of bread wheat, as well as distinct breeding status (wild progenitors, early domesticates, landraces, and cultivars) (Supplementary Table 1 and 2; for convenience, the common names of subspecies are used in this study). Notably, the collection also well represents the evolutionary trajectory of modern bread wheat[7,8,14] (Fig. 1c).

These high-coverage genomes (~6.5×) empowered high-quality calling of genetic variations in self-pollinated plants such as wheats (Supplementary Table 3). By applying the cross-ploidy variation discovery pipeline (Supplementary Fig. 2)[13], we identified ~78 million single nucleotide polymorphisms (SNPs), and constructed version 1.1 of the whole-genome

genetic variation map of wheat (VMap 1.1) (Supplementary Note 1, and Supplementary Fig. 3, and Supplementary Table 4 and 5). The false-positive error rate of variant calling, i.e., the proportion of segregating sites in the reference accession, Chinese Spring, was only 0.011%, which is similar to the error rates of high-quality SNPs in previous studies[13,22].

Spatiotemporal origin of bread wheat

Many crops transited from weedy grasses to cultivated plants through solely domestication[23]. However, for bread wheat, this early transition was coupled with an additional polyploid speciation event, from which bread wheat arose through the hybridization between tetraploid wheats (AABB) and strangulata (*Ae. tauschii* ssp. *strangulata*, DD)[24,25]. Phylogenetic analyses of VMap 1.1 corroborated two recent findings regarding the origin of bread wheat (Fig. 1b,c, and Supplementary Note 2)[13]. One is the two-stage model of wheat domestication that wild emmer (*T. turgidum* ssp. *dicoccoides*, AABB) was transformed to domesticated emmer first, then free-threshing tetraploids. The other is the identification of free-threshing tetraploid wheats as the direct donor of the AB subgenomes during the polyploid speciation of bread wheat. Although the evolutionary topology of wheat populations becomes increasingly clear, there is limited consensus on the spatiotemporal dynamics of the emergence of bread wheat[24].

As the progenitor of domesticated emmer, wild emmer comprises two subpopulations mostly confined to the northern and southern Levant in West Asia (Fig. 2a)[26,27]. Archaeological records from early Neolithic sites showed that domesticated emmer appeared in the northern Levant (Abu Hureyra and Cafer Höyük) and southern Levant (Tell Aswad) almost simultaneously ~9,800-9600 BP[28], raising a controversial question in which place emmer wheat was first domesticated[24]. By reconstructing the phylogeny of AB lineage using

150,000 random SNPs, we found that wild emmer in the northern Levant was clustered with domesticated emmer (Fig. 1b). Moreover, bread wheat showed a closer identity by state (IBS) distance with northern wild emmer rather than southern wild emmer (Fig. 2a, and Supplementary Table 13 and 14). These results support the hypothesis that emmer wheat was domesticated around the Karacadag region in the northern Levant[18,26].

The birthplace of bread wheat is also mysterious. As the distribution of wild emmer and strangulata is primarily restricted to the Levant and the south of the Caspian Sea, respectively, it was puzzling how the polyploid speciation of bread wheat could occur given the geographic isolation of parental taxa[24]. Here we identified that free-threshing tetraploids rather than wild emmer were the donor of the AB subgenomes of bread wheat (Fig. 1b and c)[13], suggesting the scenario that hexaploidization of bread wheat did not occur until free-threshing tetraploids expanded to the south of the Caspian Sea[24]. Further analyses of IBS distance showed that strangulata accessions in the southwest of the Caspian Sea have the greatest affinity to bread wheat (Fig. 2a, and Supplementary Table 15), indicating that bread wheat came into being at the southwest coast of the Caspian Sea[25].

To provide a temporal context of wheat speciation, we used SMC++[29], which combines the simplicity of sequentially Markovian coalescent and the scalability of site frequency spectrum (SFS) based approaches, to infer divergence time between wheat populations. Given the distinct evolutionary trajectories of the AB and D subgenomes of bread wheat (Fig. 1c), we inferred population split times of the AB and D lineages independently based on ~68 million neutral SNPs in VMap 1.1 (Supplementary Note 3). The results from the AB lineage showed that domesticated emmer diverged from wild emmer 10,041±160 BP, free-threshing tetraploids separated from domesticated emmer 9,269±98 BP, and bread wheat split from

135 free-threshing tetraploids 8,441±140 BP (Fig. 2b). The temporal sequence coincides nicely
136 with the oldest archaeological remains of domesticated emmer[6,28], free-threshing
137 tetraploids[30], and bread wheat[24]. Considering that hexaploidization of bread wheat involves
138 free-threshing tetraploids and strangulata simultaneously (Fig. 1c), the speciation times of
139 bread wheat inferred from the AB and D lineages should concur. However, we observed a
140 drastic gap of ~3,300 years between the two estimates, in which bread wheat diverged from
141 strangulata 11,738±112 BP (Fig. 2b).
142
143 Recent studies have identified an asymmetric wild-progenitor introgression in bread
144 wheat, where introgression is much more prevalent in the AB subgenomes (19.43%) than in
145 the D subgenome (0.49%)[13]. Given that gene flow can change the tempo of population
146 differentiation[31,32], the asymmetric introgression is likely to explain the different speciation
147 times of bread wheat inferred from AB and D subgenomes. To provide a nuanced view of
148 bread wheat speciation in the context of progenitor introgression, we investigated the
149 chronology of gene flow between wheat populations through contrasting alternative
150 demographic models[33] (Fig. 2c, and Supplementary Note 4). By comparing the observed joint
151 SFS of bread wheat and its progenitor population to the expected under a specific model, we
152 found archaic gene flow from wild emmer and domesticated emmer into bread wheat before
153 8,919 BP (95% confidence interval (CI) 8,316-9,521 BP) and 7,228 BP (95% CI 6,760-7,695
154 BP), respectively. Moreover, the best-fitting model predicted enduring and bidirectional gene
155 flow between free-threshing tetraploids and bread wheat since the emergence of bread wheat
156 ~11,700 BP. In contrast, the introgression from strangulata to the D subgenome of bread
157 wheat was more ancient, predating 9,729 BP (95% CI 9,015-10,442 BP). These results
158 suggest that the long-standing and massive gene flow to the AB subgenomes resulted in slow
159 speciation of nascent bread wheat, lasting ~3,300 years until the distinct genetic makeup of

bread wheat was established. Notably, the near-complete reproductive isolation and concomitant clean-split between bread wheat and strangulata allow the estimate of the upper time-bound of population differentiation between cultivated crops and wild relatives, which is generally intractable in diploid crops.

Trans-Eurasian dispersal of bread wheat

The spread of bread wheat across Eurasia profoundly transformed human societies[10]. To elucidate the range expansion process, we selected 225 bread wheat landraces (hereinafter referred to as landraces) from VMap 1.1 based on the accessibility of geographic information to characterize the spatiotemporal dispersal of bread wheat (Supplementary Table 18). Model-based clustering of landraces exhibited a salient east-west axis of range expansion of bread wheat originating from West Asia (Fig. 3a, and Supplementary Fig. 18), echoed by the Asian and European clades in the phylogeny of bread wheat (Fig. 1b). To reconstruct the bidirectional migration routes precisely, we applied the Estimated Effective Migration Surfaces (EEMS) method[34] to identify spatial barriers and corridors of bread wheat expansion (Fig. 3a, and Supplementary Fig. 19). EEMS presented a fast migration route westward along the northern Mediterranean coast, consistent with the uniform ancestry of landraces in the area. In contrast, EEMS eastward migration patterns identified a massive roadblock at the Pamir Mountains that splits the Inner Asian landraces into Central and South Asian populations, suggesting the further spread of bread wheat eastward through the north and south routes of the Pamir Mountains.

Landraces in East and South Asia exhibited a complex population structure, illustrating a convoluted population history of Asian bread wheat as suggested by recent archeological studies[35–39]. To disentangle the dispersal of bread wheat in the vast land of geographic and

cultural diversity, especially how bread wheat spread into China, we used qpGraph[40] to explore the relationships between local landrace populations defined by EEMS (Fig. 3b and Supplementary Table 19). By testing 61,214 candidate admixture graph models, the best-fitting graph (Z-score = -2.76) predicted three dispersal routes connecting Central and East Asia, coinciding with the postulated Southern Himalaya route[38], Hexi Corridor route[37–39], and Steppe route[35,36], respectively. The Southern Himalaya route is from Pakistan, through India, Myanmar, and Yunnan Province, into China. The mixed ancestry of landraces in southwest China (R9) provided the first evidence demonstrating the existence of the southern route[38]. The Hexi Corridor route can also be referred to as "proto-silk Road," starting from Central Asia, through the Inner Asian Mountain Corridor and Hexi Corridor to inner China. This route is the most prominent hypothesis describing wheat spread in China, verified by its abundant archaeological sites[37–39]. The Steppe route was recently proposed because the wheat remains excavated from the lower Yellow River region (~4,250 BP) are earlier than those from the upper region (~3,850 BP), indicating an alternative northern route to China via the Mongolian Steppe other than the Hexi Corridor[35]. Despite the lack of wheat samples from southern Mongolia, our results support this newly hypothesized route with genetic evidence—two populations in the lower Yellow River region (R4) and East China (R5) descended from past hybridization events (Fig. 3b), with one of the parental populations likely to be the lineage that traveled across the Mongolian Steppe. The introduction of wheat to China through the Mongolian Steppe may be related to early agropastoral societies, e.g., the Afanasievo people around the Altai Mountains, moving southward in response to the abrupt global cooling during the mid-Holocene[36].

We used SMC++[29] to calculate splitting times between locally adapted and West Asia populations to infer the timing of bread wheat dispersal across Eurasia. Given that recent

crop exchange and accompanying gene flow may reduce the divergence time estimates, we first assessed the temporal pattern of gene flow between individual local populations using *fastsimcoal2*[33]. The results showed that populations in the Iberian Peninsula, Indus Valley, Yunnan Province, and East China exhibited early gene flow to the West Asia population (Supplementary Fig. 20 and Supplementary Table 20), and thus were qualified to calculate splitting times (Fig. 3c). As these four populations probably are not strictly locally confined, we inferred the timing of bread wheat dispersal at the continental level that bread wheat may have dispersed to Europe, South Asia, and East Asia ~7,000 BP, ~6,000 BP, and ~5,400 BP, which are concordant with archeological records[35,37,38].

New *Triticum* subspecies arising from bread wheat dispersal

It becomes increasingly evident that interspecific hybridization is common during range expansion of species[31]. Bread wheat dispersal appeared to be no exception—we found several newly formed *Triticum* subspecies having their origins in sympatric hybridization between expanding bread wheat and locally preexisting tetraploid wheats. The phylogeny of *Triticum* populations showed that three hexaploid subspecies (AABBDD), including spelt (*T. aestivum* ssp. *spelta*), Macha (*T. aestivum* ssp. *macha*), and Xinjiang wheat (*T. aestivum* ssp. *petropavlovskyi*), were clustered into the tetraploid clade; similarly, a tetraploid subspecies (AABB), Persian wheat (*T. turgidum* ssp. *carthlicum*), was within the hexaploid clade (Fig. 1c and Supplementary Fig. 10). To clarify the ancestry of these outliers, we used phyloNet[41] to infer reticulate phylogenetic networks of these subspecies based on phylogenies of 9,612 orthologous genes. The result showed a mixed ancestry of the four subspecies descending from hybrids between tetraploid wheats and bread wheat, with the genetic contribution of bread wheat from 33% to 54% (Fig. 3d and Supplementary Fig. 21). It is worth noting that spelt was considered the progenitor of bread wheat because it has a primitive phenotype of

hulled seed[42], our result indicates that the phenotype is inherited from its tetraploid parent, domesticated emmer, and thus disproves the once-popular theory concerning the origin of bread wheat.

We then estimated the speciation time of the four subspecies using SMC++. To eliminate the noise from homoploid gene flow, we calculate the population splitting time between the hybrid offspring and only one of the parental taxa with different ploidy levels. The results showed spelt, Macha, Xinjiang wheat, and Persian wheat arose ~6,400 BP, 7,300 BP, ~3,300 BP, and ~6,000 BP, respectively (Fig. 3a and Supplementary Fig. 22). By calculating the IBS distance between the four subspecies and individual accessions of their parental populations, we showed that these newly formed *Triticum* subspecies likely originate from Europe, West Asia, and Central Asia (Fig. 3a and Supplementary Fig. 23-26).

Genetic heritage of bread wheat expansion

The trans-Eurasian dispersal of bread wheat may have involved extensive adaptive changes in the genome while colonizing novel environments. To investigate how the adaptation process affects the genetic diversity of bread wheat, we examined the correlations between SNPs and environmental variables of 225 landraces using redundancy analysis (RDA)[43] (Supplementary Table 22). These environmental variables include altitude and 19 bioclimatic variables related to either temperature or precipitation. We found that these variables explained 13.44% of the total SNP variance. To evaluate the confounding effect of environmental adaptation and isolation-by-distance, we performed a similar RDA analysis using latitude and longitude, instead, as explanatory variables, finding that only 6.05% SNP variance was explained (Supplementary Fig. 27). The results demonstrate the importance of environmental factors in shaping the adaptive genetic diversity of bread wheat. By

conducting individual RDA analyses on environmental variable categories, temperature-related variables (adjust r^2 = 0.11) exhibited larger SNP variance than did precipitation (adjust r^2 = 0.075) and altitude (adjust r^2 = 0.013) (Fig. 4a). However, in search for the most important environmental variables, precipitation of the warmest quarter appeared on the top of the list (Fig. 4b), suggesting the complexity of local adaptation of bread wheat. To investigate the regional heterogeneity of adaptation, we performed RDA analyses on environmental variable categories using landraces from West Asia (WA), Europe (EU), Inner Asia (IA), East Asia (EA) and Southern Himalaya (SH) (Supplementary Fig. 28). The result showed that environment variables in WA explain the least SNP variance compared with other regions. In addition, the relative proportions of SNP variance explained by temperature, precipitation, and altitude varied in the five regions (Fig. 4c, Supplementary Fig. 29 and Supplementary Table 23). The results indicate that the accumulation of adaptive alleles from the range expansion has shaped a diverse adaptation landscape of bread wheat.

To identify genomic regions associated with adaptation, we performed cross-population composite likelihood ratio (XP-CLR)[44] analyses to detect selective sweeps between paired populations from the five populations mentioned above. Collectively, 185,865 selective sweeps were discovered under the top 5% XP-CLR score threshold. As these sweeps may stem from selections of human preference, farming practices, etc., we then conducted environmental association analyses using Bayenv[45] to narrow down the candidate sweep regions to those related to environmental factors. Based on associations between 20 environmental variables and allele frequency of 1.5M SNPs in 13 populations (Supplementary Fig. 30-32 and Supplementary Table 24), the analysis identified 269,279 adaptation-associated SNPs (top 5% Bayes factor) intersecting with selective sweeps from XP-CLR, with an average of 2.15-fold enrichment for coexisting with sweep regions

(Supplementary Fig. 33 and Supplementary Table 25). A total of 19,999 genes were identified as being involved in the environmental adaptation of bread wheat, including 123 cloned genes that regulate critical agronomic traits, such as disease resistance and abiotic stress response, etc. (Supplementary Fig. 34, and Supplementary Table 26-28), indicating the value of adaptation-associated genes in improving agronomic traits of modern wheat.

Convergent adaptation of bread wheat to early flowering across Eurasia

To further characterize these adaptation-associated genes, we focused on genes relevant to flowering time because flowering time is agriculturally important for crops and commonly deemed the most critical trait determining plant adaptation[46]. Remarkably, we found that the gene *Ppd-D1* exhibited convergent adaptation to early flowering and best showcased local adaptation of bread wheat. *Ppd-D1* on chromosome 2D is the primary determinant of photoperiod response in bread wheat. Dysfunctional *Ppd-D1* exhibits photoperiod insensitivity and early flowering phenotypes, which is crucial to the adaptation of bread wheat to global environments[47]. A total of three loss-of-function alleles of *Ppd-D1* have been identified so far in wheat populations (Fig. 4d), two of which are causal genetic variants, including a ~2kb deletion at upstream, and a 5-bp deletion in gene exons[47,48]. The XP-CLR analysis between IA and SH populations identified selective footprints on *Ppd-D1*, predicting an increased frequency of causative alleles of *Ppd-D1* in the SH population (Fig. 4e). However, the two causative alleles did not exist in SH landraces (Fig. 4f and Supplementary Fig. 35). Instead, we found a novel stop-gain mutation of *Ppd-D1* in the SH landraces, particularly enriched in the population from the Tibetan Plateau (Fig. 4f and g). We speculated that the stop-gain allele helped adapt bread wheat to the short growing season in high-altitude and low-temperature areas (>3,000 m). To test the hypothesis, we divided the SH landraces into high-altitude and low-altitude subpopulations and performed XP-CLR

analysis (Supplementary Table 29). The result showed that the XP-CLR score (99.75% quantile) on *Ppd-D1* became more significant when compared with the score (97.61% quantile) between IA and SH landraces, indicating *Ppd-D1* is involved in high-altitude adaptation (Fig. 4e and Supplementary Fig. 36). Furthermore, we found a strong correlation between the allele frequency of the stop-gain mutation and the average altitude of subpopulations from SH landraces ($r^2 = 0.778$, Fig. 4h), showing the causal effect of the stop-gain mutation in the high-altitude adaptation of bread wheat. Taken together, the three causative alleles of *Ppd-D1* complement each other in geographic distribution, with the stop-gain mutation in South Asia, the ~2kb deletion in East Asia, and the 5-bp deletion in Europe (Fig. 4g, i, and j), illustrating a highly diverse but convergent adaptation of bread wheat across Eurasia through its changing flowering time.

Population size fluctuation of wheats

Compared with cultivars, crop wild relatives has received relatively little attention from the evolutionary perspective[4]. To decipher the population dynamics of wild wheats, we reconstructed the history of effective population size (N_e) of *Triticum-Aegilops* species using SMC++[29]. We found that N_e of *Aegilops* subspecies, including strangulata, tauschii (*Ae. tauschii* ssp. *tauschii*, DD), and speltoilds (*Ae*. speltoides, BB/SS), appeared to decline constantly in the last 100 thousand years (Supplementary Fig. 37). In contrast, all the *Triticum* species experienced a marked population size expansion during the Holocene. For bread wheat and early domesticates, such as domesticated einkorn, domesticated emmer, and free-threshing tetraploids, the population growth may reflect the cultivation history of these populations[6]; whereas for wild wheats that had never been domesticated, such as wild einkorn (*T. monococcum* ssp. *aegilopoides*, AA), urartu (*T. urartu*, AA), and wild emmer (Fig. 1 and Supplementary Fig. 37, 38), such population growth may result from their mixed growing

335 with early domesticates for thousands of years[24,49]. Strikingly, we found a ubiquitous
336 population contraction right after the population growth for all the *Triticum* populations
337 except for modern cultivars of bread wheat. The population decline occurred sequentially
338 with ploidy levels—the diploids came first, then the tetraploids, and lastly, the hexaploid
339 landraces. Intriguingly, the rise and fall of N_e of diploids, tetraploids, and hexaploids
340 complemented each other in the Holocene timeline even without impact from drastic climate
341 change of glacial periods (Supplementary Fig. 37). Archeological studies showed that
342 domesticated einkorn and domesticated emmer thrived since the Neolithic Age until they
343 were gradually replaced by durum wheat (*T. turgidum* ssp. *durum*, AABB), spelt, and bread
344 wheat during the Bronze Age (~5,000 BP - ~3,000 BP)[6,24]. The N_e fluctuation of wheats
345 coincides with the shifts of human food choice from einkorn and emmer wheat to bread wheat.
346 Despite largely being a natural evolutionary process, the population size decline of wild
347 wheats is disturbing—the N_e of diploids and tetraploids in *Triticum* was reduced by 81.70%
348 in the past two thousand years (Fig. 5a).
349
350 Rapid climate change is likely to impact the biodiversity of wheats profoundly[50]. To
351 evaluate the adaptive capacity of *Triticum-Aegilops* species, we conducted biogeographical
352 modeling to predict the response of wheats to the future climate. For bread wheat, we used a
353 tree-based machine learning approach, gradient forest, to model allele frequency of genome-
354 wide SNPs from 13 populations (Supplementary Fig. 30 and 39) with 19 bioclimatic
355 variables. The adaptation-associated SNPs identified previously (Supplementary Fig. 33)
356 presented a much faster turnover rate along environmental gradient than the randomly chosen
357 SNPs (Supplementary Fig. 40). Using the adaptation-associated SNPs, we predicted the shift
358 of allele frequency, namely genetic offset, of local landraces between present and future
359 climates during 2040-2060 and 2080-2100. Local bread wheat populations showed varying

360 degrees of genetic offset, with the highest value appearing in regions of the Indus Valley and
361 Inner Asia, indicating that wheat production in the two regions is the most vulnerable to
362 climate change (Fig. 5b, c and Supplementary Fig. 41). Since we did not have a large enough
363 sample size to model allele frequency of individual wild wheat populations, we used Species
364 Distribution Modeling (SDM)[51] to predict the future habitats of wild wheats. Overall, we
365 observed either a contraction of wild wheats' habitats or shifting of their geographical ranges
366 to the north (Supplementary Fig. 42-45). As such, two of the critical progenitors of bread
367 wheat, wild emmer and strangulata (Fig. 1c), clearly showed the projected change of species
368 distribution (Fig. 5d). It is worth noting that wild emmer, which is the ultimate source of
369 genetic diversity of bread wheat[52], may become a threatened species requiring conservation
370 in a few decades.
371

372 ## Discussion

373 The changing climate is threatening global food security[53]. The evolution of major crops
374 through immense space and time provides an unparalleled opportunity to dissect the
375 environmental adaptation of plants and further help address the climate challenge. The
376 population history of bread wheat and its wild relatives has long been controversial[6,21,24]. By
377 leveraging a comprehensive set of genomes from the genera *Triticum* and *Aegilops*, we
378 systematically unraveled the spatiotemporal history of bread wheat and its wild relatives in
379 the Holocene, in particular the origin and range expansion of bread wheat, population size
380 dynamics of wild wheats, and hybridization events between the two groups. We also found
381 the high-altitude adaption of Tibetan landraces through a stop-gain allele of an essential
382 flowering time gene (*Ppd-D1*). Remarkably, this allele and the other two independent loss-
383 of-function alleles of *Ppd-D1* showed a pattern of convergent adaptation of bread wheat to
384 early flowering across Eurasia, indicating the important role of evolutionary constraint in

385 shaping adaptive landscape of bread wheat. Meanwhile, the adaptive diversity demonstrates
386 wild wheats as an invaluable resource providing pre-adapted alleles for wheat breeding.
387 However, some of the most important wild populations exhibited a disturbing population
388 decline driven by shifts of human food choice and environmental change, illustrating the
389 pressing need to protect wild wheats. Taken together, our work of reconstructing the
390 population history of *Triticum-Aegilops* species has laid an essential foundation to dissect
391 the genetics of wheat adaptation effectively. It also provides a research paradigm to explore
392 population history and adaptive genetic diversity for all crops. This study will facilitate well-
393 informed efforts in protecting wheat biodiversity and breeding climate-resilient crops in the
394 future.

Online Methods

Sampling of wheat accessions

A total of 795 accessions of bread wheat and its wild relatives (a.k.a. wheats) from 73 countries were used in this study (Supplementary Table 1). This natural population contains 6 species and 25 subspecies, covering all subspecies with AA, AABB, and AABBDD genome types in *Triticum*, as well as BB/SS, and DD genome types in *Aegilops*. We integrated 414 accessions from the VMap1.0[13], 92 accessions from Northwest A&F University[18], and 244 accessions sequenced at China Agricultural University[17]. In addition, to collect all possible genetic donors of bread wheat, we newly sequenced 50 additional accessions, including 10 accessions of speltoids (*Ae. speltoides* ssp. *speltoides*), 10 accessions of spelt (*T. aestivum* ssp. *spelta*), 3 accessions of synthesis hexaploid wheat, and 27 accessions of strangulata (*Ae. tauschii* ssp. *strangulata*) in this study (Supplementary Table 2 and 3). Plant materials are available at the Chinese Crop Germplasm Resources Information System (CGRIS), National Small Grains Collection (NSGC), and Genebank Gatersleben of Leibniz Institute of Plant Genetics and Crop Plant Research (IPK).

Sequencing, reads mapping, and genetic variation discovery

Newly added 50 samples were sequenced in two batches. The first batch of 27 accessions was sequenced with the BGISEQ500 machine using 100-base-pair pair-end reads. The second batch of 23 accessions was sequenced with the MGISEQ-2000RS machine using 150-base-pair pair-end reads (Supplementary Table 3). Reads with more than half of bases with a quality value of less than Q20 or more than 5% of 'N' bases were removed from the raw data. Given the different ploidy levels and genome types across the population, we applied a custom pipeline for genetic variation calling[13] (Supplementary Fig. 2). The reference genome (IWGSC RefSeq v1.0)[12] was divided into five taxonomic groups based on their genome types,

420 namely AA, BB/SS, AABB, AABBDD, and DD taxa, respectively. Then, all accessions were
421 mapped to their corresponding reference with BWA-MEM[54]. GATK[55] was used to obtain the
422 raw SNPs for each genome type. We used HaplotypeCaller of GATK with parameters of "--
423 native-pair-hmm-threads 5 -L -R -I -O -ERC GVCF". The SNP filtering procedures are the
424 same as those applied in our previous work[13].
425
426 **Construction of VMap 1.1 by merging SNPs from five genome types**
427 The taxa containing A, B, and D lineages have different coalescence times. When
428 coalescence is deep, the reference bias of SNP calling can be severe. For a fair comparison
429 between A, B, and D lineages, only SNPs in syntenic sites were retained in the final genetic
430 variation dataset, as previously described[13]. Finally, we built the variation library of VMap
431 1.1 (genetic variation map of wheat version 1.1), containing ~78M SNPs (Supplementary
432 Note. 1). Then, we genotyped 795 accessions by scanning their bam files using HapScanner
433 (https://github.com/PlantGeneticsLab/TIGER/wiki/HapScanner)[13]. The SNPs of VMap 1.1
434 are summarized in Supplementary Table 5.
435
436 **Reconstructing gene tree chronograms for genera *Triticum* and *Aegilops***
437 (1) Identification of orthologous genes across *Triticum* and *Aegilops*. *Hordeum vulgare*[56],
438 *Triticum urartu*[57], *Aegilops tauschii*[58], wild emmer[27], durum[59], Chinese Spring[12] were
439 downloaded from ensemble plants (http://plants.ensembl.org/index.html). Ortholog
440 prediction was carried out using SwiftOrtho[60] with default parameters. Results of SwiftOrtho
441 were parsed into three groups of A, B, and D subgenomes. Finally, 4,971 reciprocal genes
442 were found to estimate the phylogenetic history of the A, B, and D lineages.
443 (2) Local assembly of orthologous genes. As part of species within genera *Triticum* and
444 *Aegilops* do not have reference genomes (e.g., einkorn and speltoides) that challenges

phylogeny reconstruction at the species level, we performed local assembly of orthologous genes for individual *Triticum-Aegilops* subspecies. First, fastq sequences of orthologous genes were extracted from bam files of 25 subspecies. Second, we assembled the fastq sequences for each gene of 25 subspecies with SRAssembler[61]. Third, the assembled segments were anchored to the reference genome of bread wheat[12]. Last, we used Muscle[62] to perform multiple sequence alignments, and only the sequence alignments longer than 1,000 bp were kept in the final assembly-based gene sequences (4,806 reciprocal genes).

(3) Reconstruction of the species tree. We analyzed ortholog genes of each subspecies jointly in a coalescent analysis to clarify the relationship of the *Triticum* and *Aegilops* subspecies. Multispecies coalescent analysis was carried out in BEAST2[63] to calculate gene tree topology. All gene trees were rooted with *Hordeum vulgare* with a secondary normally distributed calibration in the root about 15 Mya[7]. The analysis was conducted under the HKY substitution model for each gene. MCMC chains were run for 50 million generations with parameters sampled every 10,000 generations. Analyses were examined for convergence in Tracer v1.7.1[63], and a burn-in of 25 million generations was discarded. Those trees with mean posterior probability across all nodes of ≥0.85 were kept using TreeAnnotatorv.2.5.1[63]. Finally, the global tree was made ultrametric using the chronos function in the ape R package (http://ape-package.ird.fr/) and drawn using ggtree[64] in R.

Population phylogenetic analysis

We selected 150,000 random SNPs from corresponding subgenomes for each phylogeny reconstruction. Phylogenetic trees of each group were reconstructed using RAxML[65] software with 100 bootstrap replicates in the GTRGAMMA model, and the output tree was plotted in iTOL[66]. We reconstructed the tree using a 100-times bootstrap with barley as the outgroup for the A, B, AB, or D lineage. We reconstructed the tree using 100 bootstrap

replicates for the AB lineages with wild emmer as the outgroup. In addition, according to the outgroup of the hexaploid individuals closest to strangulata in the tree of D lineage, a tree of D subgenome in ABD genome type was constructed with 100 bootstrap replicates. The parameters of RAxML were "-f a -m GTRGAMMA -p 12,346 -x 12,346 -# 100."

Population genetic differentiation statistics

(1) Genetic distance between populations. Pairwise distances between the bread wheat and its progenitors in the whole-genome level were calculated as the fraction differences between pairs of samples for each individual using PLINK[67] with the formulation: 1-IBS, where IBS was identity by state. (2) F_{ST} values. We estimated F_{ST} values for populations using vcftools[68], and Weir and Cockerham's calculation in 1 Mb non-overlapping windows. Pairwise comparisons between each subspecies in each lineage were calculated.

Speciation time estimation

We used SMC++ v1.15.4 to reconstruct effective population size histories for each paired population separately[29]. To mitigate the effect of selection on the estimate of the most recent common ancestor, we filtered out SNP sites (1) within genes; (2) outside the possible regulate elements (3kb upstream and downstream of the gene). Next, VCF files containing 15 pseudo-diploid genotypes was generated from randomly selected 30 individuals, as the previous study for the self-fertilization plants[69]. We then partitioned VCF files into SMC haploblock files for each pair, and partitioned each chromosome by using the vcf2smc function in SMC++. Subsequently, we used a polarization error of 0.5 and the mutation rate of 6.5×10^{-9} in the ESTIMATE function of SMC++ to estimate past effective population sizes[70,71] (*Ne*). Results were scaled to real-time by applying a generation time of 1 year and plotted on a linear timescale.

In addition, we used the effective size dynamics of two subpopulations, calculating split times between subpopulations in a cross-coalescent framework of SMC++ "OMP_NUM_THREADS=1 smc++ split". Finally, confidence intervals were estimated from population divergence time based on 20 resampling replicates (Supplementary Fig. 12 and Supplementary Note. 3).

Model inference of demographic history

We applied the site frequency spectrum (SFS) to infer demographic scenarios using coalescent simulations to approximate the likelihood of a given model. Demographic parameter estimation was implemented in *fastsimcoal* version 2.6[33].

(1) Data preparation and processing. To estimate migration rates among AB lineages and D lineage, we generated the observed SFS file as the input of *fastsimcoal2* through the following steps: First, identifying the ancestral allele. *Hordeum vulgare*[56] and *Aegilops tauschii*[58] are the ancestral groups for AB lineages, while the *Hordeum vulgare* and *Triticum urartu*[57] are the same ancestral group for D lineage. The NUCmer program implemented in the latest release of MUMmer4[72] was used to align the genomes of the outgroups to that of Chinese Spring[12] with "--maxmatch -g 1,000 -c 90 -l 40". Sites overlapping with VMap 1.1 were retained to infer ancestral alleles. For biallelic SNPs, alleles identical in two outgroups were identified. Second, generating the site-frequency spectrum (SFS). In AB lineages, 20 accessions were randomly selected from each subpopulation and paired subpopulations (wild emmer, domesticated emmer, free-threshing tetraploids, and bread wheat AB) produced 2D-SFS. In D lineage, 20 accessions were randomly selected from two subpopulations (stranglulata and bread wheat D) that produced 2D-SFS. Here, we used easySFS (https://github.com/isaacovercast/easySFS) to convert the VCF to various SFS formats.

519 (2) Model selection and model fitting. Five demographic models (early Geneflow, no gene
520 flow, recent gene flow, different gene flow matrices, and constant gene flow) used the site
521 frequency spectrum (SFS) to fit model parameters to the observed data by performing
522 coalescent simulations. For each model, the fit to the observed SFS was maximized using the
523 composite-likelihood method implemented in *fastsimcoal2* with the following options: "-N
524 100,000 -L 50", with other options by default. We used wide search ranges with log-uniform
525 distributions for all parameter estimates and assumed a generation time of 1 year and a
526 constant mutation rate of 6.5×10^{-9} mutation/generation/site. Subsequently, we performed
527 100 independent *fastsimcoal2* runs for each demographic model to determine the parameter
528 estimates leading to the maximum likelihood. We compared different gene flow scenarios in
529 AB and D lineage to get the best-fitting model.
530 (3) Bootstrap analysis. We estimated confidence intervals for the model with maximum
531 likelihood by estimating parameters on 50 bootstrap data sets. The bootstrap data sets were
532 obtained by randomly re-sampling 20 accessions in specific subpopulations to match the
533 original data set size. Then, for each bootstrapped dataset, we obtained SFS with easySFS
534 software. Next, re-estimated parameters using the same settings as the original data set, but
535 with 20 replicate runs instead of 100, due to computational constraints. To obtain the 95%
536 confidence intervals, we calculated the 2.5% and 97.5% percentiles of the estimate
537 distribution obtained with R.
538
539 **Population structure projected on the map**
540 We used population structure to provide insights into the migratory patterns of the bread
541 wheat landrace. SNPs in D lineage were selected for ADMIXTURE[73] by applying the
542 following criteria: (1) SNPs with linkage disequilibrium (LD) above 0.2 were removed using
543 Plink "--indep-pairwise 50 10 0.2", and (2) SNPs with MAF ≥0.05. Geographical projections

544 of population structure were obtained using the hclust function in cluster package
545 (https://cran.r-project.org/web/packages/cluster/index.html) in R. The spatial prediction was
546 based on a Gaussian model, which supposes that the covariance matrix is stationary. We
547 implemented the map projection by mapPie function in rworldmap package (https://cran.r-
548 project.org/web/packages/rworldmap/) in R.
549
550 **Estimated effective migration surfaces**
551 Estimated effective migration surfaces (EEMS) is an approach to estimate genetic migration
552 patterns according to a given geographic region[34]. We computed genetic dissimilarity
553 matrices using EEMS and assigned geographical coordinates to each sample from each
554 district to contrast geographic and genetic distances between demes. Migration surface
555 contours were estimated using 800 demes for all sections of Eurasia. We ran MCMC analysis
556 for 500,000 MCMC iterations, including 300,000 burn-in iterations, and repeated the process
557 with different seeds to ensure the convergence of the MCMC chains with parameters by
558 default.
559 Final spatial visualizations illustrating migratory surfaces were generated using R
560 scripts provided by EEMS.PLOT function from the rEEMSplots package
561 (https://github.com/dipetkov/eems). To test the robustness of the models, we applied a jack-
562 knife sampling approach and repeated the EEMS runs after iteratively excluding isolates
563 from a single district.
564
565 **Admixture graph modeling for landrace subpopulations in East and South Asia**
566 We sought to find explicit population history models that can infer the dispersal routes in
567 East Asia. Therefore, we reconstructed admixture graphs[74] for Asian landrace subpopulations
568 defined by the EEMS classification[34]. Bread wheat landraces in East and South Asia were

569 divided into ten subpopulations (R1-R10), and the individual list is available in
570 Supplementary Table 19. First, we filter the data set using the following criteria: SNPs with
571 linkage disequilibrium (LD) above 0.2 were removed using Plink "--indep-pairwise 50 10
572 0.2", no more than 5% missing data. Second, the CONVERTF function from AdmixTools[40]
573 was used to produce eigenstrat format data files, and the qpgraph function was used to
574 evaluate whether graph models fit the data, using the West Asian population as an outgroup.
575 We then computed f_2-, f_3- and f_4-statistics measuring allele sharing of two, three, or four
576 sets of subpopulations and reported the maximum $|Z|$-score between predicted and observed
577 values. To explore the space of all possible admixture graphs, we used a heuristic search
578 algorithm named qpbrute[75]. Given an outgroup with which to root the graph, a stepwise
579 addition order algorithm was used for adding leaf nodes to the graph. At each step, the
580 insertion of a new node was tested at all branches of the graph, except the outgroup branch
581 (West Asia landrace). Where a node could not be inserted without producing f_4 outliers
582 ($|Z| \geq 3$), then all possible admixture combinations were also attempted. If a node could not
583 be inserted via either approach, that sub-graph was discarded. If the node was successfully
584 inserted, the remaining nodes were recursively inserted into that graph. All possible starting
585 node orders were attempted to ensure complete coverage of the graph space.
586 The effective use of qpGraph is to determine the relationships between subpopulations
587 when the relationships indicated by phylogenetic trees are unclear[76]. We could put known
588 subgroup relationships into the graph to reduce the amount of computation. We construct the
589 admixture graphs of the South Asian and East Asian groups to fix the known structures using
590 the heuristic search method, respectively. Then, the same heuristic algorithm was used to
591 build the admixture graphs of 10 subpopulations. Finally, we fitted 61,214 possible
592 admixture graph models and recorded the three graphs that left no f4 outliers ($|Z| < 3$). We

593 then found the best-fit graph using the admixturegraph package[77] in R to compute the
594 marginal likelihood of these three models and their Bayes Factors (BF).
595
596 **Phylogenetic-network analysis**
597 PhyloNet[41] inferred species hybridization events using the proportions of gene tree
598 topologies to locate past hybridization within a phylogeny in the presence of incomplete
599 lineage sorting. We inferred the hybridization events in *Triticum* through the following three
600 steps. Firstly, get the individual ortholog gene trees. RAxML[65] was used to build an ML gene
601 tree for each identified ortholog gene (n = 4,806) under the GTRGAMMA substitution model.
602 Subsequently, species networks modeled incomplete lineage sorting and gene flow using a
603 pseudo-maximum likelihood approach were carried out with PHYLONET v.3.6.1[41] with the
604 command "InferNetwork_MPL" and using the individual gene trees. Finally, network
605 searches were performed using only nodes in the rooted ML gene trees with bootstrap support
606 of at least 75%, allowing for 0–4 reticulations and optimizing the branch lengths and
607 inheritance probabilities of the returned species networks under pseudo-likelihood.
608
609 **Abiotic variables collection and redundancy analyses**
610 We downloaded climate-related variables data and altitude information from WorldClim
611 (https://www.worldclim.org/), which provides monthly climate precipitation and
612 temperature data at 30 seconds (~1 km^2) resolution for the period 1,970-2,000. Then the
613 EXTRACT function of R package RASTER v.3.3.13 (https://cran.r-
614 project.org/web/packages/raster) was used for geographic data analysis. These climate
615 variables included eleven temperature variables (Annual Mean Temperature, Mean Diurnal
616 Range (Mean of monthly (max temp - min temp)), Isothermality (Temp2/Temp7) (×100),
617 Temperature Seasonality (standard deviation ×100), Max Temperature of Warmest Month,

618 Min Temperature of Coldest Month, Temperature Annual Range (Temp5-Temp6), Mean
619 Temperature of Wettest Quarter, Mean Temperature of Driest Quarter, Mean Temperature
620 of Warmest Quarter, Mean Temperature of Coldest Quarter) and eight precipitation variables
621 (Annual Precipitation, Precipitation of Wettest Month, Precipitation of Driest Month,
622 Precipitation Seasonality (Coefficient of Variation), Precipitation of Wettest Quarter,
623 Precipitation of Driest Quarter, Precipitation of Warmest Quarter, Precipitation of Coldest
624 Quarter).

625 We used redundancy analysis (RDA)[43] to identify multiple climate variables important
626 for explaining SNP variance in bread wheat landraces. We ran RDA utilizing a subset of 20K
627 randomly chosen SNPs with no missing and MAF >0.05 for response variables. Then, RDA
628 with variance partitioning was conducted to quantify the proportion of genome-wide SNP
629 variation explained by 20 abiotic categories variables. To identify abiotic variables associated
630 with genome-wide divergence among different regions, we conducted RDA using three
631 significant variables (temperature, precipitation, and altitude) for different regions, including
632 West Asia (WA), Europe (EU), Inner Asia (IA), East Asia (EA) and Southern Himalaya (SH)
633 groups. Finally, all RDAs were conducted using the R package VEGAN
634 (https://github.com/vegandevs/vegan).

635

636 **Selective sweeps detection in different regions of bread wheat**

637 The XP-CLR statistic[44] was used to identify selective sweeps in different regions of bread
638 wheat landraces. We divided Eurasian bread wheat landraces into five subgroups (West Asia,
639 Europe, Inner Asia, East Asia, and Southern Himalaya) based on geographical, genetic, and
640 ecological differences and calculated the selective sweep for each pair of subgroups. XP-
641 CLR was run with the grid size of 10 kb, the maximum number of SNPs of 500 within a
642 window, and the correlation level as 0.95. The genetic distance was estimated from the

643 recombination rate data from a previous publication[12]. The R package GenWin
644 (https://cran.r-project.org/web/packages/GenWin) was used to normalize XP-CLR statistics
645 and detect the boundary of genomic regions with smoothness = 2,000 and method = 4. We
646 considered the top 5% of the statistic results from each population as the threshold under
647 selective sweep and calculated different thresholds for different subgenomes.
648
649 **Environmental association analysis in bread wheat**
650 Association between local environment and SNP frequency was identified using Bayenv
651 2.0[45]. 20 environmental variables (11 temperature variables, 8 precipitation variables, and
652 altitude) were obtained from the WorldClim (https://worldclim.org/). A total of 13 wheat
653 populations was identified on the basis of geographic and environmental variables using the
654 k-means approach implemented in R package cluster (https://cran.r-
655 project.org/web/packages/cluster). To control for population structure, we used LD
656 independent SNPs of A, B and D lineage ("--indep-pairwise 50 10 0.2") to estimate the
657 covariance matrix of 13 populations with 100,000 iterations. The association between the 1.5
658 million SNPs and the 20 environmental variables was tested with 10,000 iterations for each
659 SNP. The median value of the Bayes factor was calculated for each SNP using data from five
660 independent Bayenv runs. The top 5% value of Bayes factor was set as the threshold to select
661 environment associated SNPs for each of the 20 environmental variables.
662
663 **Detecting adaptation-associated SNPs and genes**
664 The regions with the top 5% XP-CLR score and environment-associated SNPs were
665 identified as adaptation-associated regions. Genes in such regions were considered candidate
666 adaptation-associated genes. Next, we used the snpEff (version 5.0)[78] to annotate genetic

667 variations and predict the functional effects of SNPs in the longest transcript of adaptation-
668 associated genes, based on the gene annotation of IWGSC gtf v1.1[12].

669 For the *Ppd-D1* gene, known functional 5 bp deletion (33,953,310) and 16 bp insertion
670 (33,952,522) loci genotypes were extracted from the raw VMap1.1 which contains the indels
671 variants. The upstream ~2kb deletion was identified by profiling the reads depth in whole-
672 genome sequencing data for each sample.

673

674 **Population size fluctuation over time**

675 We used SMC++[29] to infer historical effective population sizes of bread wheat and its wild
676 relatives. The SNP data pre-processing was as same as inferring speciation time of bread
677 wheat. Both gene and regulatory regions were removed, and pseudo-diploid genotypes were
678 generated for the self-fertilization plants[69] (Supplementary Note. 3). We set the upper bound
679 of the number of generations as 100,000 and the lower bound to 100. We set the number of
680 spline knots used in the internal representation of population size history to 30. The
681 estimation process assumed a mutation rate of 6.5×10^{-9} mutation/generation/site. All other
682 parameters were set to the default values. The final representation of the history of effective
683 population size was made using a generation time of one year. To estimate the variance of
684 effective population size in SMC++, we resampled the pseudo-diploid genotypes for each
685 subpopulation (n = 20). If the number of subspecies is less than 5, we repeat the coalescent
686 process of the same samples for 20 times. Finally, we plotted all independent SMC++
687 analyses for each considered subpopulation. The combined result was drawn in R using the
688 "stat_smooth" function, and the confidence interval level is 0.95.

689

690 **Estimating the genetic offset of bread wheat in the future climates**

Biogeographical modeling was used to identify environmental factors important to allele frequency change and to detect how allele frequency shifts along that factors[79]. We divided 225 bread wheat landraces into 13 populations as described previously. We tested two sets of SNPs for modeling using gradient forest[80]. One SNP set was created by randomly choosing 30,000 SNPs from 225 samples. The other was a random selection of 30,000 SNPs from the adaptation-associated SNPs described previously.

We extended the gradient forest analysis to predict "genetic offset (GO)"[79]. Here, "genomic offset" measures the mismatch between current genotype and projected genotype in the future environment using associations across current environment gradients as a baseline[81,82]. Based on the NorESM1-M Global Climate Model (GCM), the future climate variables of 2050 (2040-2060) and 2090 (2080-2100) under four different greenhouse gas scenarios, Representative Concentration Pathways (RCPs), including RCP2.6, RCP4.5, RCP6.0, and RCP8.5, were retrieved from WorldClim (https://worldclim.org/). Those four RCPs represent different gas emission scenarios, reflecting conditions ranging from moderate (RCP2.6) to the extreme (RCP8.5)[83]. Nineteen bioclimatic variables from both current and predicted climates were transformed for each raster based on importance in predicting genomic variation using the gradient forest model. The Euclidean distance between the current and projected future values is the genetic offset of individual populations. The genetic offset was predicted by package gradientForest (https://gradientforest.r-forge.r-project.org/) and projected by package rasterVis (https://cran.r-project.org/web/packages/rasterVis/index.html) in R.

Species distribution modeling

Species distribution models (SDMs) combine observations of species distribution with environmental estimates[51]. We implemented correlative species distribution using

716 geographic coordinates of bread wheat and its wild relatives retrieved from online germplasm
717 databases, such as the U.S. National Plant Germplasm System (https://npgsweb.ars-
718 grin.gov/gringlobal/search) to construct species distribution models (SDM)[84,85]. Nineteen
719 environmental predictors retrieved from WorldClim (https://worldclim.org/) were used in
720 our final SDM modeling. SDMs were then generated with package dismo in R (https://cran.r-
721 project.org/web/packages/dismo/index.html) with three modeling algorithms, Generalized
722 Additive Models, Generalized Linear Models, and Random Forests. The species occurrence
723 data were combined with 50 pseudo-absence data that were randomly generated within the
724 area of study. Models were trained using 70% of data and tested with the remaining 30%.
725 Each modeling algorithm was run 100 times and was evaluated via true skill statistics (TSS).
726 The final models were then used for each species to project the potential distribution of each
727 species under both current and projected future climatic (2040-2060 and 2080-2100)
728 conditions.

729 **References:**

730 1. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield Trends Are Insufficient
731 to Double Global Crop Production by 2050. *PLoS One* **8**, e66428 (2013).
732 2. Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D. & Schroeder, J. I.
733 Genetic strategies for improving crop yields. *Nature* **575**, 109–118 (2019).
734 3. Hickey, L. T. *et al.* Breeding crops to feed 10 billion. *Nat. Biotechnol.* **37**, 744–754
735 (2019).
736 4. Castañeda-Álvarez, N. P. *et al.* Global conservation priorities for crop wild relatives.
737 *Nat. Plants* **2**, 1–6 (2016).
738 5. Shiferaw, B. *et al.* Crops that feed the world 10. Past successes and future challenges
739 to the role played by wheat in global food security. *Food Secur.* **5**, 291–317 (2013).
740 6. Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R. & Martin, W. Genetics and
741 geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* **3**, 429–441
742 (2002).
743 7. Marcussen, T. *et al.* Ancient hybridizations among the ancestral genomes of bread
744 wheat. *Science* **345**, 1250092 (2014).
745 8. Glémin, S. *et al.* Pervasive hybridizations in the history of wheat relatives. *Sci. Adv.*
746 **5**, eaav9188 (2019).
747 9. Pont, C. *et al.* Tracing the ancestry of modern bread wheats. *Nat. Genet.* **51**, 905–911
748 (2019).
749 10. Diamond, J. M. *Guns, germs, and steel : the fates of human societies*. (New York :
750 Norton, [2005]).
751 11. Harari, Y. N. author. *Sapiens : a brief history of humankind*. (First U.S. edition. New
752 York : Harper, [2015]).
753 12. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in

754 wheat research and breeding using a fully annotated reference genome. *Science* **361**,
755 eaar7191 (2018).
756 13. Zhou, Y. *et al.* Triticum population sequencing provides insights into wheat adaptation.
757 *Nat. Genet.* **52**, 1412–1422 (2020).
758 14. Zhou, Y. *et al.* Introgressing the Aegilops tauschii genome into wheat as a basis for
759 cereal improvement. *Nat. Plants* **7**, 774–786 (2021).
760 15. Balfourier, F. *et al.* Worldwide phylogeography and history of wheat genetic diversity.
761 *Sci. Adv.* **5**, eaav0536 (2019).
762 16. He, F. *et al.* Exome sequencing highlights the role of wild-relative introgression in
763 shaping the adaptive landscape of the wheat genome. *Nat. Genet.* **51**, 896–904 (2019).
764 17. Guo, W. *et al.* Origin and adaptation to high altitude of Tibetan semi-wild wheat. *Nat.*
765 *Commun.* **11**, 1–12 (2020).
766 18. Cheng, H. *et al.* Frequent intra- and inter-species introgression shapes the landscape
767 of genetic variation in bread wheat. *Genome Biol.* **20**, 136 (2019).
768 19. Gaurav, K. *et al.* Population genomic analysis of Aegilops tauschii identifies targets
769 for bread wheat improvement. *Nat. Biotechnol.* (2021).
770 20. Matsuoka, Y. Evolution of Polyploid Triticum Wheats under Cultivation: The Role of
771 Domestication, Natural Hybridization and Allopolyploid Speciation in their
772 Diversification. *Plant Cell Physiol.* **52**, 750–764 (2011).
773 21. Zhao, X., Fu, X., Yin, C. & Lu, F. Wheat speciation and adaptation: perspectives from
774 reticulate evolution. *aBIOTECH* **2**, 386–402 (2021).
775 22. Ramu, P. *et al.* Cassava haplotype map highlights fixation of deleterious mutations
776 during clonal propagation. *Nat. Genet.* **49**, 959–963 (2017).
777 23. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: Genetics of
778 domestication and diversification. *Nature Reviews Genetics* **14**, 840–852 (2013).

779 24. Faris, J. D. Wheat domestication: Key to agricultural revolutions past and future. in
780 *Genomics of Plant Genetic Resources: Volume 1. Managing, Sequencing and Mining*
781 *Genetic Resources* (eds. Tuberosa, R., Graner, A. & Frison, E.) 439–464 (Springer
782 Netherlands, 2014).
783 25. Wang, J. *et al.* Aegilops tauschii single nucleotide polymorphisms shed light on the
784 origins of wheat D-genome genetic diversity and pinpoint the geographic origin of
785 hexaploid wheat. *New Phytol.* **198**, 925–937 (2013).
786 26. Luo, M.-C. *et al.* The structure of wild and domesticated emmer wheat populations,
787 gene flow between them, and the site of emmer domestication. *Theor. Appl. Genet.*
788 **114**, 947–959 (2007).
789 27. Avni, R. *et al.* Wild emmer genome architecture and diversity elucidate wheat
790 evolution and domestication. *Science* **357**, 93–97 (2017).
791 28. Nesbitt, M. & Samuel, D. From staple crop to extinction? The archaeology and history
792 of the hulled wheats. *Hulled wheats. Proc. First Int. Work. Hulled Wheats* **4**, 41–100
793 (1996).
794 29. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population
795 history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
796 30. Araus, J. L., Slafer, G. A., Romagosa, I. & Molist, M. FOCUS: Estimated wheat yields
797 during the emergence of agriculture based on the carbon isotope discrimination of
798 grains: Evidence from a 10th millenuim BP site on the euphrates. *J. Archaeol. Sci.* **28**,
799 341–350 (2001).
800 31. Purugganan, M. D. Evolutionary Insights into the Nature of Plant Domestication. *Curr.*
801 *Biol.* **29**, R705–R714 (2019).
802 32. Edelman, N. B. *et al.* Genomic architecture and introgression shape a butterfly
803 radiation. *Science* **366**, 594–599 (2019).

804 33. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust
805 demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
806 34. Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure
807 with estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2015).
808 35. Zhou, X. *et al.* 5,200-year-old cereal grains from the eastern Altai Mountains redate
809 the trans-Eurasian crop exchange. *Nat. Plants* **6**, 78–87 (2020).
810 36. Long, T. *et al.* The early history of wheat in China from 14C dating and Bayesian
811 chronological modelling. *Nat. Plants* **4**, 272–279 (2018).
812 37. Liu, X. *et al.* From ecological opportunism to multi-cropping: Mapping food
813 globalisation in prehistory. *Quat. Sci. Rev.* **206**, 21–28 (2019).
814 38. Stevens, C. J. *et al.* Between China and South Asia: A Middle Asian corridor of crop
815 dispersal and agricultural innovation in the Bronze Age. *The Holocene* **26**, 1541–1555
816 (2016).
817 39. Betts, A., Jia, P. W. & Dodson, J. The origins of wheat in China and potential pathways
818 for its introduction: A review. *Quat. Int.* **348**, 158–168 (2014).
819 40. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093
820 (2012).
821 41. Than, C., Ruths, D. & Nakhleh, L. PhyloNet: A software package for analyzing and
822 reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**, 1–16
823 (2008).
824 42. Mac Key, J. Species relationship in Triticum. *Hereditas* **Suppl. 2**, 237–275 (1966).
825 43. van den Wollenberg, A. L. Redundancy analysis an alternative for canonical
826 correlation analysis. *Psychom. 1977 422* **42**, 207–219 (1977).
827 44. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective
828 sweeps. *Genome Res.* **20**, 393–402 (2010).

829 45. Günther, T. & Coop, G. Robust identification of local adaptation from allele
830 frequencies. *Genetics* **195**, 205–220 (2013).
831 46. Liang, Y., Liu, H. J., Yan, J. & Tian, F. Natural Variation in Crops: Realized
832 Understanding, Continuing Promise. *Annu. Rev. Plant Biol.* **72**, 357–385 (2021).
833 47. Beales, J., Turner, A., Griffiths, S., Snape, J. W. & Laurie, D. A. A Pseudo-Response
834 Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat
835 (Triticum aestivum L.). *Theor. Appl. Genet.* **115**, 721–733 (2007).
836 48. Guo, Z., Song, Y., Zhou, R., Ren, Z. & Jia, J. Discovery, evaluation and distribution
837 of haplotypes of the wheat Ppd-D1 gene. *New Phytol.* **185**, 841–851 (2010).
838 49. Tanno, K. I. & Willcox, G. How fast was wild wheat domesticated? *Science* **311**, 1886
839 (2006).
840 50. Scheffers, B. R. *et al.* The broad footprint of climate change from genes to biomes to
841 people. *Science (New York, N.Y.)* **354**, aaf7671 (2016).
842 51. Elith, J. & Leathwick, J. R. Species distribution models: Ecological explanation and
843 prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).
844 52. M., S. A. *et al.* Precipitation drives global variation in natural selection. *Science* **355**,
845 959–962 (2017).
846 53. Anderson, R., Bayer, P. E. & Edwards, D. Climate change and the need for agricultural
847 adaptation. *Current Opinion in Plant Biology* **56**, 197–202 (2020).
848 54. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
849 MEM. *arXiv* **00**, 1–3 (2013).
850 55. Aaron McKenna *et al.* The Genome Analysis Toolkit: A MapReduce framework for
851 analyzing next-generation DNA sequencing data. *Genome Res.* 254–260 (2009).
852 56. Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the
853 barley genome. *Nature* **544**, 427–433 (2017).

854 57. Ling, H. Q. *et al.* Genome sequence of the progenitor of wheat A subgenome Triticum
855 urartu. *Nature* **557**, 424–428 (2018).
856 58. Luo, M. C. *et al.* Genome sequence of the progenitor of the wheat D genome Aegilops
857 tauschii. *Nature* **551**, 498–502 (2017).
858 59. Maccaferri, M. *et al.* Durum wheat genome highlights past domestication signatures
859 and future improvement targets. *Nat. Genet.* **51**, 885–895 (2019).
860 60. Hu, X. & Friedberg, I. SwiftOrtho: A fast, memory-efficient, multiple genome
861 orthology classifier. *Gigascience* **8**, 1–12 (2019).
862 61. McCarthy, T. W., Chou, H. C. & Brendel, V. P. SRAssembler: Selective Recursive
863 local Assembly of homologous genomic regions. *BMC Bioinformatics* **20**, 1–13
864 (2019).
865 62. Edgar, R. C. MUSCLE: A multiple sequence alignment method with reduced time and
866 space complexity. *BMC Bioinformatics* **5**, 1–19 (2004).
867 63. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics
868 with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
869 64. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R Package for
870 Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other
871 Associated Data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
872 65. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis
873 of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
874 66. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display
875 and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245
876 (2016).
877 67. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-
878 based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

879 68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–
880 2158 (2011).
881 69. Cubry, P. *et al.* The Rise and Fall of African Rice Cultivation Revealed by Analysis
882 of 246 New Genomes. *Curr. Biol.* **28**, 2274-2282.e6 (2018).
883 70. Gutaker, R. M. *et al.* Genomic history and ecology of the geographic spread of rice.
884 *Nat. Plants* **6**, 492–502 (2020).
885 71. Zeng, X. *et al.* Origin and evolution of qingke barley in Tibet. *Nat. Commun.* **9**, 1–11
886 (2018).
887 72. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS*
888 *Comput. Biol.* **14**, e1005944 (2018).
889 73. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry
890 in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
891 74. Peter, B. M. Admixture, Population Structure and F-statistics. *Genetics* **202**, 1485–
892 1501 (2016).
893 75. Máire Ní, L. *et al.* The Evolutionary History of Dogs in the Americas. *Science* **85**, 81–
894 85 (2018).
895 76. Lipson, M. Applying f4-statistics and admixture graphs: Theory and examples. *Mol.*
896 *Ecol. Resour.* **20**, 1658–1667 (2020).
897 77. Leppälä, K., Nielsen, S. V. & Mailund, T. Admixturegraph: An R package for
898 admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017).
899 78. Cingolani, P. *et al.* A program for annotating and predicting the effects of single
900 nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster
901 strain w1118; iso-2; iso-3. *Fly (Austin).* **6**, 80–92 (2012).
902 79. Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level
903 modelling of biodiversity: Mapping the genomic landscape of current and future

environmental adaptation. *Ecol. Lett.* **18**, 1–16 (2015).

80. Ellis, N., Smith, S. J. & Roland Pitcher, C. Gradient forests: Calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).

81. Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y. & Keller, S. R. Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Mol. Ecol. Resour.* **21**, 2749–2765 (2021).

82. Fitzpatrick, M. C., Keller, S. R. & Lotterhos, K. E. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **361**, 83–86 (2018).

83. van Vuuren, D. P. *et al.* The representative concentration pathways: An overview. *Clim. Change* **109**, 5–31 (2011).

84. Aguirre-Liguori, J. A., Ramírez-Barahona, S. & Gaut, B. S. The evolutionary genomics of species' responses to climate change. *Nat. Ecol. Evol.* **5**, 1350–1360 (2021).

85. Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N. & Zimmermann, N. E. Uncertainty in ensembles of global biodiversity scenarios. *Nat. Commun.* **10**, 1–9 (2019).

Acknowledgements:

We thank Yalong Guo, and Song Ge (Institute of Botany, Chinese Academy of Sciences) for their suggestions. This work is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24020201), the National Natural Science Foundation of China (31921005 and 31970631), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24040102).

Author contributions:

X.Z. and Y.G. performed data analysis, plotted manuscript figures, and drafted the manuscript; Y.L., C.Y. and J. W. collected plant materials; L.P., A.B., X.S., D.X., Z.Z., J.Z., J.X., X.Y., S.X., M.Z, and P.K. helped with data analysis. X.F. and Z.L. contributed to project coordination. F.L. conceived the idea, coordinated the project, and wrote the final manuscript. All authors discussed the results and commented on the manuscript.

Competing interests:

The authors declare no competing interests.

Data availability statement:

The raw sequence data were deposited in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/) under the accession number of PRJCA005979. The genotype data from VMap 1.1 are publicly available at the Genome Variation Map (https://bigd.big.ac.cn/gvm) under accession number GVM000272.

Code availability statement

947 The custom code for demographic history is available at

948 https://github.com/xuebozhao16/VMap1.1-Population_history_of_wheats

Figure legends:

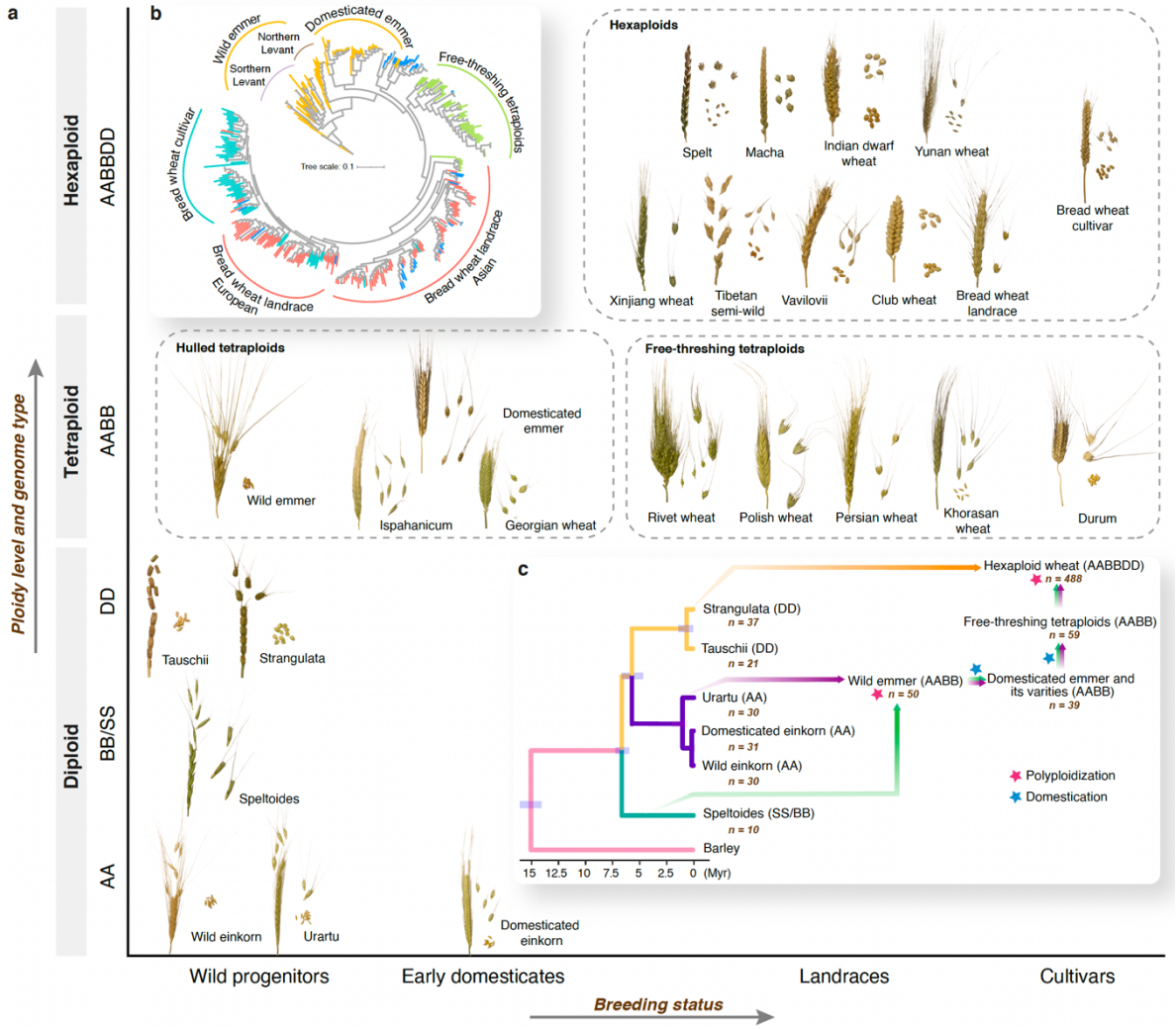


Fig. 1 | The representative collection of wheats in this study. a, Common name, spike morphology, ploidy level, genome type, and breeding status of wheat accessions. **b,** Relationship of wheat accessions in the AB lineage illustrated by the phylogeny tree with wild emmer as the outgroup. **c,** Evolutionary relationship of bread wheat and its wild progenitors in the genera *Triticum* and *Aegilops*. The sample size of the individual taxa was labeled. The chronogram of the phylogeny was obtained by calculating the divergence of orthologous genes between species (Methods and Supplementary Fig. 9).

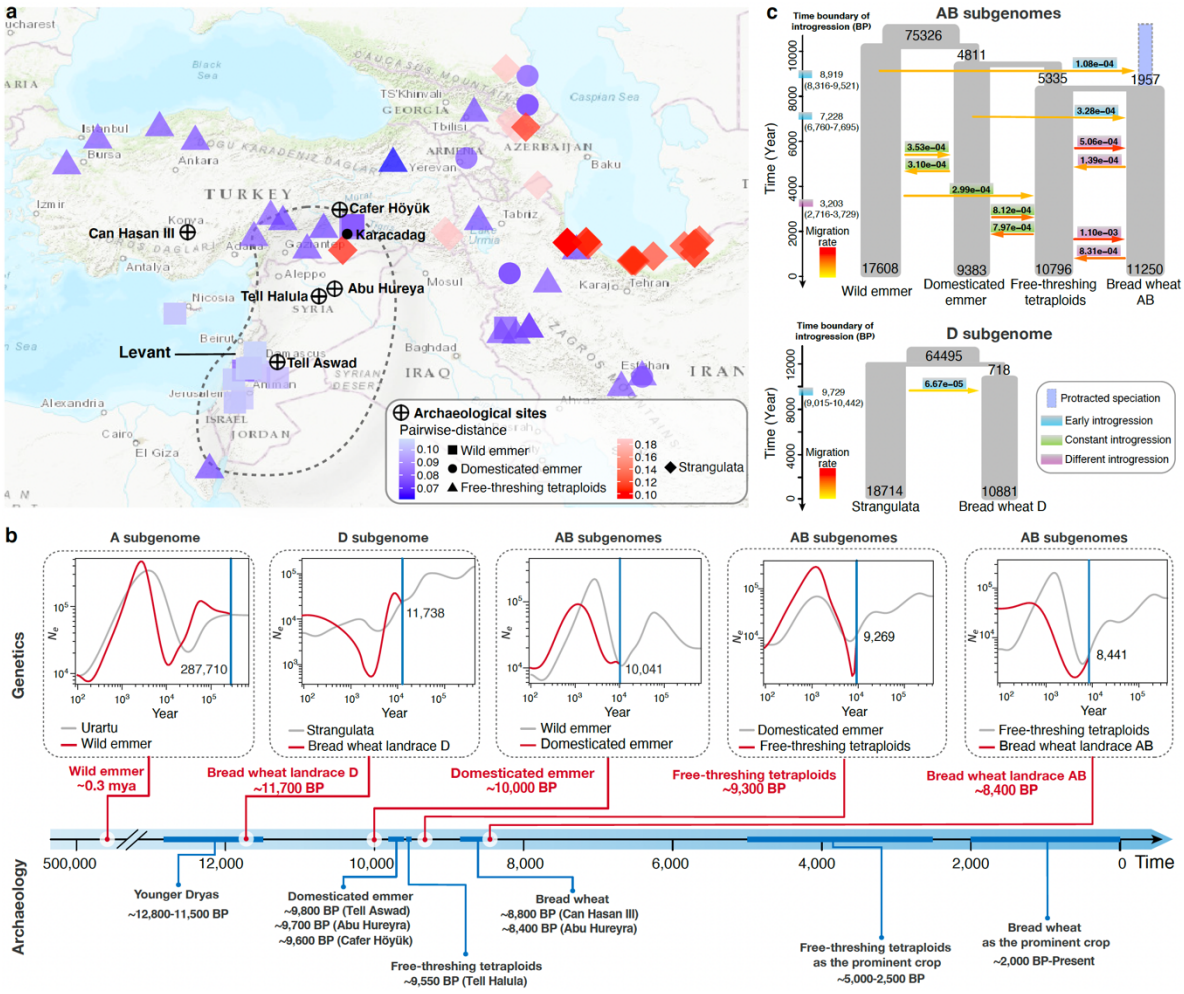
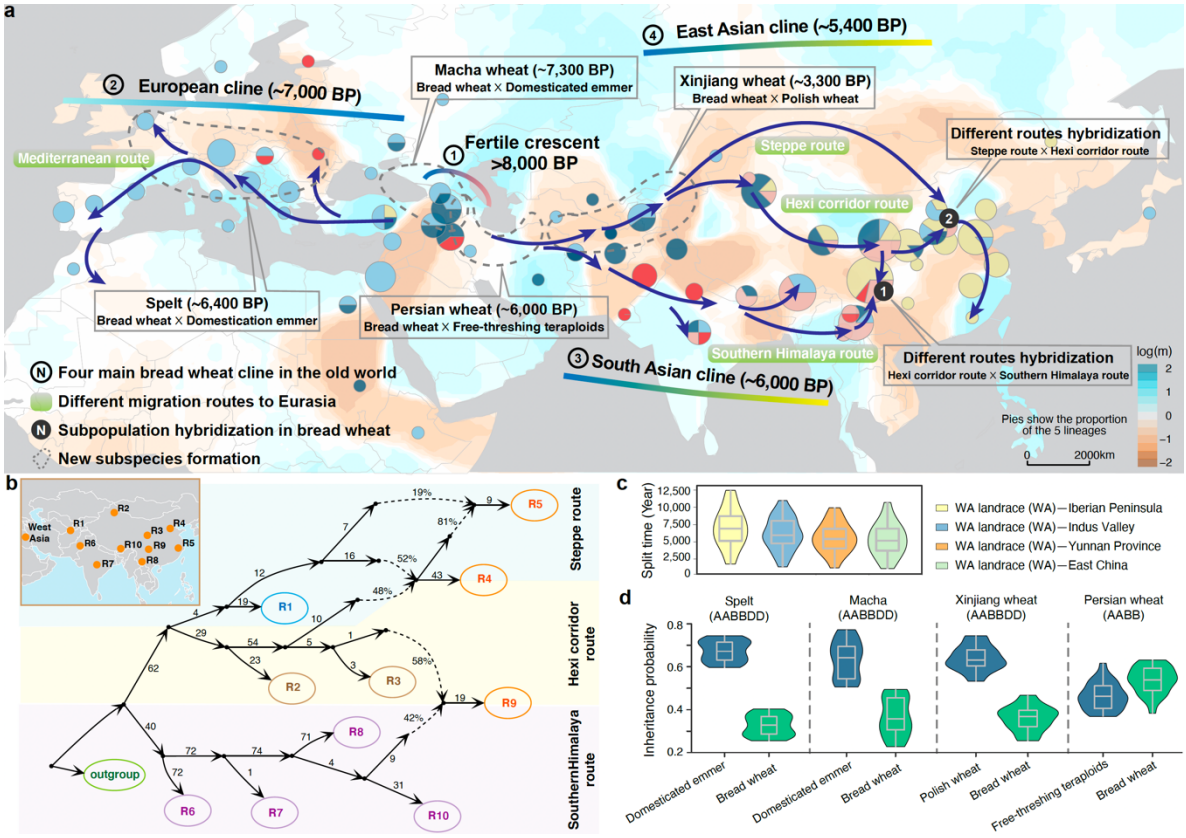


Fig. 2 | Demographic models of the bread wheat speciation. a, A geographic affiliation of IBS distances across bread wheat and its progenitors. Color scale indicates the distance of the AB subgenomes (blue) and the D subgenome (red) between bread wheat and progenitors. The map was created using the R package rworldmap. **b,** Timeline of evolutionary events related to bread wheat speciation. The top is the timeline of population split between wheats inferred from SMC++. The bottom is the wheat evolutionary timeline derived from archaeological evidence. **c,** The best supported demographic model of the speciation and introgression in wheats for AB subgenomes and D subgenome. The width of each grey rectangle indicates the estimated effective population size (N_e). Arrows among the grey rectangles are the migration rates (m) among different populations, and only $2N_em$ >1 is shown. The colored rectangle at the timeline indicates the time boundary of introgression.



971

972 **Fig. 3 | Trans-Eurasian expansion of bread wheat. a,** Proposed dispersal routes of bread
973 wheat in Eurasia. The map colors showed the estimated effective migration surfaces (EEMS)
974 representing migration barriers (orange) and channels (cyan). Pies on the map showed the
975 ancestral proportion of the five lineages. Arrows were the estimated migration routes from
976 the Fertile Crescent to Europe and Asia. Boxes mark subpopulation hybridization and new
977 subspecies formation events, and the stippled areas represent the regions where the
978 hybridization events took place. **b,** Admixture graph model identifies the hybridization
979 events of bread wheat in ten regions along the eastward route. Solid lines with arrowheads
980 represent uniform ancestries, and attached numbers show scaled drift parameter f_2. Dashed
981 lines represent mixed ancestries, and attached values indicate estimated proportion of
982 ancestry. **c,** Distribution of split times estimated from cross-coalescence analysis of different
983 regions. The median and quartiles with whiskers reaching up to 1.5 times the interquartile
984 range are shown in boxplots. **d,** Inheritance probability of four *Triticum* subspecies formed
985 through hybridization during bread wheat dispersal.

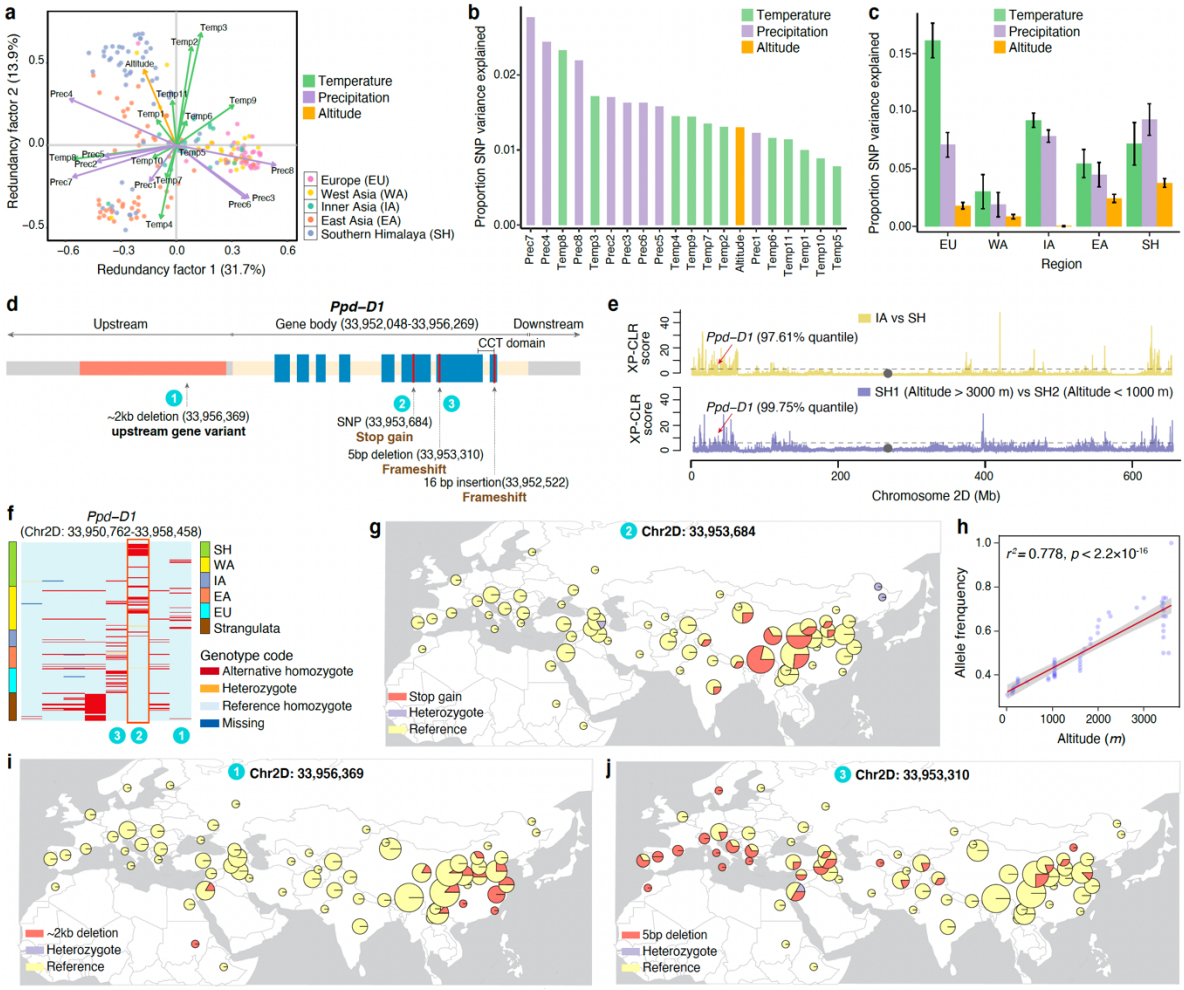


Fig. 4 | Geographic expansion reshaped the adaptive genetic diversity of bread wheat. **a,** Landraces mapped on the first two canonical axes of Redundancy analysis (RDA). Arrows represent 20 environmental factors (11 temperature factors, 8 precipitation factors, and altitude) that are correlated with genotype of landraces. Colored points representing accessions from different regions: Europe (EU), West Asia (WA), Inner Asia (IA), East Asia (EA), and South Himalayas (SH). **b,** Ranked importance of environmental factors based on individual RDA analyses. **c**, Proportion of total SNP variance explained in RDA by environmental variable categories in each region. **d,** Sequence *Ppd-D1* gene on the chromosome 2D of the reference genome (Chinese Spring). Three causative loss-of-function alleles and non-causative frameshift mutation are marked with red rectangles. The light-yellow rectangle represents the gene body. Blue rectangles represent exons. **e,** Selective

sweeps on chromosome 2D to identify adaptive footprints on *Ppd-D1*. Top: IA vs. SH. Bottom: SH1 (Altitude > 3000 m) vs. SH2 (Altitude < 1000 m). The horizontal dotted lines indicate the top 5% genome-wide cut-off level. Arrows marked the position and top quantile of the *Ppd-D1* gene. **f,** Haplotypes of *Ppd-D1* gene in strangulata and bread wheat landrace. The numbers represent three loss-of-function genetic variants corresponding to **d**. The colored bars on the left represente different species/populations. **g,** Geographic distribution of the stop-gain mutation (number 2) of *Ppd-D1* gene. **h,** Correlation between frequency and altitude of stop-gain mutation (number 2) of *Ppd-D1* gene. **i,** Geographic distribution of ~2kb deletion (number 1) of *Ppd-D1* gene. **j,** Geographic distribution of 5-bp deletion (number 3) of *Ppd-D1* gene. Orange indicates the proportion of three loss-of-function haplotypes in **g**, **i** and **j**, respectively. Geographic maps in **g**, **i** and **j** were created using the R package rworldmap.

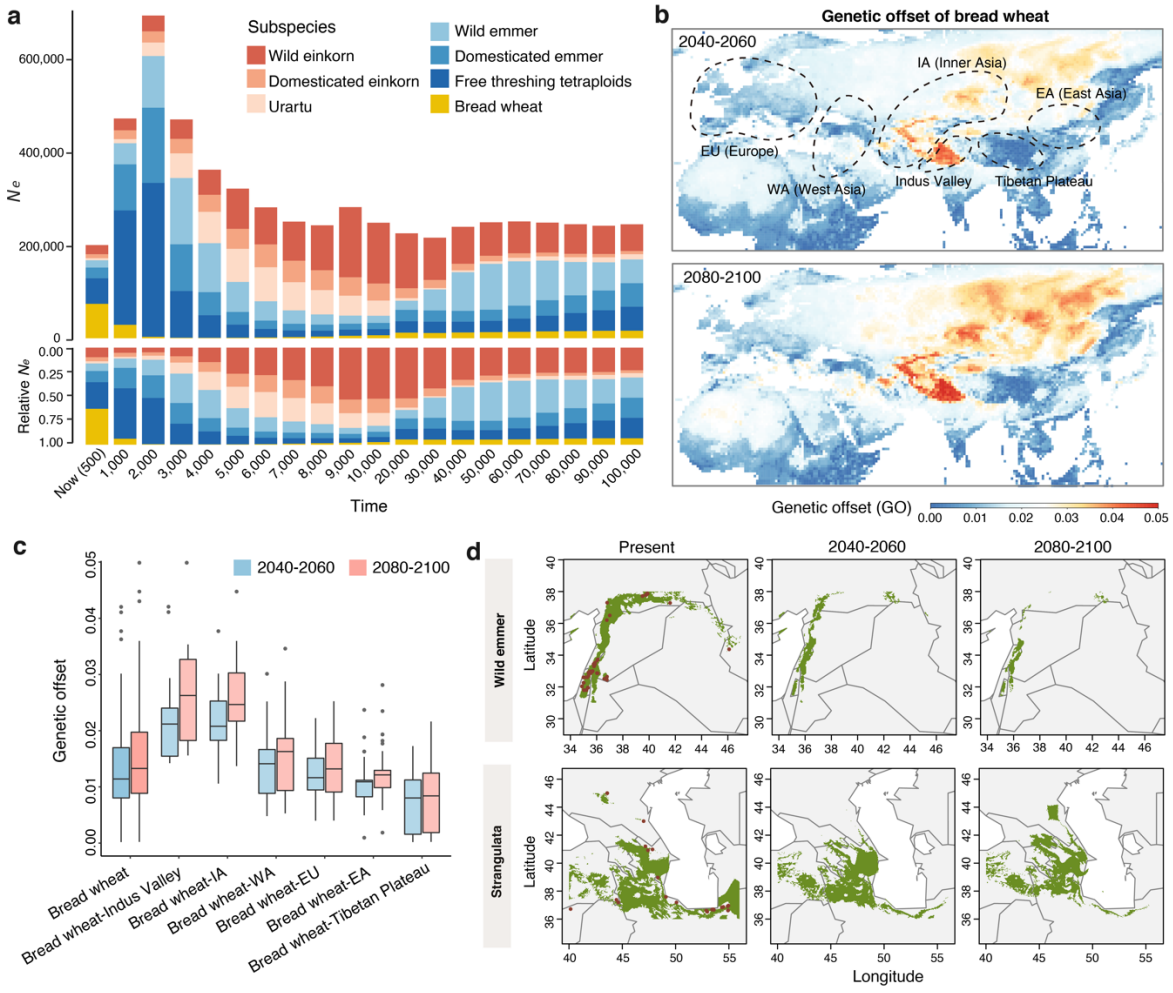


Fig. 5 | The population size fluctuation of wheats from the past to the future. a, Holocene population dynamics in wheats. The top of this figure depicts the *Ne* for seven populations, and the bottom of this figure is the relative *Ne* proportion of each population. **b,** Genetic offset (GO) of bread wheat landrace based on 2040-2060 RCP8.5 and 2080-2100 RCP8.5 projections. **c,** Genetic offset of bread wheat landrace in six geographical regions, corresponding to **b**. The median and quartiles with whiskers reaching up to 1.5 times the interquartile range are shown in boxplots. **d,** Species distribution models (SDMs) projected the geographical range of wild emmer and strangulata populations in the present and future (2040-2060 and 2080-2100). Red dots pointed to the location of the samples in VMap1.1 and the USDA website (https://npgsweb.ars-grin.gov/gringlobal/search). The green shaded areas are suitable predicted regions for planting.