the plant journal

SEB Society for Experimental Biology

The Plant Journal (2019) **99,** 201–215 doi: 10.1111/tpj.14421

Phylogenetic and population structural inference from genomic ancestry maintained in present-day common wheat Chinese landraces

Xiyan Wu[1,†], Baoxu Ding[2,†], Bingqi Zhang[2], Jiaojiao Feng[2], Yibing Wang[2], Chao Ning[1], Haidan Wu[2], Fan Zhang[1], Qun Zhang[3], Ning Li[2], Zhibin Zhang[2], Xuhan Sun[2], Quanchao Zhang[3], Wenying Li[4], Bao Liu[2], Yinqiu Cui[1,3,*] and Lei Gong[2,*]

[1]*School of Life Sciences, Jilin University, Changchun 130012, People's Republic of China,*

[2]*Key Laboratory of Molecular Epigenetics of the Ministry of Education, Northeast Normal University, Changchun 130024, People's Republic of China,*

[3]*Research Center for Chinese Frontier Archaeology, Jilin University, Changchun 130012, People's Republic of China, and*

[4]*Xinjiang Cultural Relics and Archaeology Institute, Ürümchi 830000, PR China*

Received 19 December 2018; accepted 10 May 2019; published online 28 May 2019.

*For correspondence (e-mails cuiyq@jlu.edu.cn; gongl100@nenu.edu.cn).

†These authors contributed equally to this work.

SUMMARY

Hexaploid common wheat is one of the most important food crops worldwide. Common wheat domestication began in the Fertile Crescent of the Near East approximately 10 000 years ago and then spread west into Europe and eastward into East Asia and China. However, the possible spreading route into and within China is still unclear. In this study, we successfully extracted DNA from single ancient wheat seeds and sequenced the whole genome of seven ancient samples from Xiaohe and Gumugou cemeteries in Xinjiang, China. Genomic inference and morphological observation confirmed their identity as hexaploid common wheat grown in prehistoric China at least 3200 years before present (BP). Phylogenetic and admixture analyses with RNA-seq data of modern hexaploid wheat cultivars from both China and Western countries demonstrated a close kinship of the ancient wheat to extant common wheat landraces in southwestern China. The highly similar allelic frequencies in modern landraces of the Qinghai-Tibetan plateau with the ancient wheat support the previously suggested southwestern spreading route into highland China. A subsequent dispersal route from the Qinghai-Tibetan plateau margins to the Yangtze valley was proposed in this study. Furthermore, the common wheat populations grown in the Middle and Lower Yangtze valley wheat zones were also proposed to be established by population admixture with the wheat grown in the Upper Yangtze valley. Our study reports ancient common wheat sequences at a genome-wide scale, providing important information on the origin, dispersal, and genetic improvement under cultivation of present-day wheat landraces grown in China.

Keywords: ancient DNA, dispersal route, hexaploid common wheat, genomic ancestry, population structure.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit https://doi.org/10.1111/tpj.14453.

INTRODUCTION

Common wheat (*Triticum aestivum*, genome BBAADD) originated from two sequential allopolyploidization events, involving the initial allotetraploidization between two diploid species, *T. urartu* (AA) and a yet undetermined goatgrass species closely related to *Aegilops speltoides* (BB), and followed by allohexaploidization between a primitively domesticated tetraploid wheat of *T. turgidum* (BBAA) and *Aegilops tauschii* (DD) (Dvorák, 1976, Feldman, 1976; Huang *et al.*, 2002; Dvorák and Akhunov, 2005). As a milestone in the evolution of human civilization, common wheat originated approximately 10 000 years ago in the Fertile Crescent and then spread westward into Europe and eastward into East Asia (Salamini *et al.*, 2002). In the eastward spreading of cultivation into East Asia, including into China approximately 4500 years ago, the limited gene flow from *Ae. tauschii* (DD) and *T. turgidum* subsp. *dicoccoides* (BBAA) is believed to have conserved the population genetic structure of the original hexaploid common

wheat (Dvorák *et al.*, 2006; Zhao, 2009; Wang *et al.*, 2013; Zhou *et al.*, 2018).

Therefore, research into the genetic structure and diversity of Chinese common wheat populations may provide valuable clues about how the wheat population evolved, spread, and further domesticated in China. Previous studies have mainly focused on populations of modern wheat landraces to characterize and compare the genomic population structure and diversity within extant wheat landraces, which are locally adapted to different wheat-growing zones in China (Cavanagh *et al.*, 2013; Wurschum *et al.*, 2013; Zhou *et al.*, 2018). These studies have highlighted the adaptive evolutionary history of wheat cultivation and its possible entry into and further dispersal within China. Notably, it has been established that precise inferences about the population evolution of interested species necessitate the analysis of ancient archeological samples (Gugerli *et al.*, 2005; Ramos-Madrigal *et al.*, 2016; Schaefer *et al.*, 2016). Facilitated by recent excavations of more ancient samples and advances in whole-genome surveys of ancient genomes, studies on population evolution in various crops, such as barley, maize, grape, *Gossypium*, have integrated modern cultivars, wild accessions and ancient samples (Palmer *et al.*, 2012; Mascher *et al.*, 2016; Ramos-Madrigal *et al.*, 2016; Wales *et al.*, 2016). These results have provided more accurate and insightful information into the origin, ancient genomic ancestry, initial domestication and subsequent dispersal of modern crops within certain geographical regions. Nevertheless, studies on common wheat population evolution in China using ancient wheat DNA have not been reported. Therefore, without data from ancient DNA, those inferred entry sites and following dispersal routes remain contentious (Zhou *et al.*, 2018).

Recently, evolutionary analyses of ancient wheat DNA have been carried with samples excavated in Europe and China (Li *et al.*, 2011; Bilgic *et al.*, 2016). These pioneering studies have mainly focused on the possible genomic identity (ploidy level) of ancient samples, the age of the sampled ancient wheat DNA, and/or the evolution of specific DNA regions in ancient wheat samples (Li *et al.*, 2011; Bilgic *et al.*, 2016). A unique whole-genomic analysis was completed on sedimentary ancient wheat DNA extracted from paleosol sediment samples excavated at Bouldnor Cliff in UK (Smith *et al.*, 2015). However, because the data of these ancient wheat samples were generated by mixed sampling and sequencing, they are not suitable for further population studies (Weiss *et al.*, 2015). Xinjiang, as a major geographic intersection between the West and the East, may have played an important role in the eastward spreading of wheat (Millward, 2007; Li *et al.*, 2011). Abundant, well preserved cereal remains, including ancient wheat seeds, have been excavated in several prehistoric cemeteries, such as Xiaohe and Gumugou cemeteries, located in the Taklimakan Desert, and they were determined to have similar radiocarbon dates and archeological context (Xie *et al.*, 2013; Li *et al.*, 2015). These ancient wheat seeds are a valuable resource for completing individual DNA sequencing to reveal the population evolutionary history of wheat since its entry into China especially for its further spreading within China.

In this study, we initially characterized the morphological features of ancient wheat seeds excavated from the cemeteries in Xinjiang. The identity of the ancient wheat seeds as hexaploid common wheat was further confirmed using our improved whole-genome sequencing technique for the ancient wheat DNA extracted from each single ancient wheat seed. The well maintained ancient genomic ancestry and the similarity of its population structure with those of the landraces currently grown in southwestern China were also characterized. There are indications that the population evolution of common wheat might be underpinned by both natural and human-mediated selection processes. The inferred similarities in allelic frequency between ancient and modern wheat landraces provide clues about the exact wheat dispersal route after being introduced to China. Overall, our study provides the ancient wheat sequences at a genome-wide scale, representing the most detailed exploration to date of wheat evolution at the population level through the integration of both modern and ancient wheat genome sequences. Our study provides important insights into the origin, dispersal, and domestication of present-day wheat landraces grown in China.

RESULTS

Morphological similarities in assayed modern common and ancient wheat seeds

Prior to genomic analyses, we surveyed the morphological characteristics of the ancient wheat seeds relative to seeds of the contemporary tetraploid wheats with the BBAA genome (including wild emmer wheat, *Tritium turgidum* subsp. *dicoccoides* and domesticated emmer *durum*, and *T. carthlicum*), hexaploid common wheat with the BBAADD genome (*T. aestivum*), and their diploid progenitors, containing the A-genome, *T. urartu* and *T. monococcum*, the B (S)-genome, *Ae. speltoides* and *Ae. sharonensis*, and the D-genome, *Ae. tauschii* (Figure 1). All ancient and modern wheat seeds were photographed by a super-depth-of-field three-dimensional microscope at ×30 magnification (Experimental procedures). For this analysis, seven dehusked and well preserved ancient wheat seeds from each of the different tombs at the Xiaohe and Gumugou cemeteries were randomly selected and morphologically examined (Figure S1). In addition to cursory observations of the oblong/ovate shape, detailed measurements were made for each seed at three subdivided sections, which included the upper

embryo region, middle endosperm crease region, and bottom tip region (Figure 1). Results showed that: (i) the ancient wheat seeds from the Xiaohe and Gumugou cemeteries in general had morphological features highly similar to those of hexaploid common wheat, with clear contribution from *Ae. tauschii* (Table S1); (ii) within the upper embryo region, the germs of ancient wheat displayed similar wrinkle patterns as those in modern cultivated tetraploid and hexaploid wheat varieties but significantly differed from other diploid and tetraploid seeds (Figure 1a); (iii) around the ventral crease region, the ancient wheat seeds displayed a unique conspicuous crease that was present in BB/SS, DD, and BBAADD species but was absent from AA, BBAA wild, and BBAA cultivated species (Figure 1b); and (iv) at the bottom tips, long brushy hairs were observed in ancient wheat seeds, AA, BBAA wild, and BBAADD species (Figure 1c). Collectively, the ancient wheat seeds displayed overall similarity in every morphological feature with those of hexaploid common wheat, indicating that they were hexaploid wheat with the BBAADD genome.

High endogenous rate of sufficient authentic ancient wheat DNA sequences generated by single-seed library construction

DNA extraction from archeological, especially paleobotanical, plant specimens and further high-throughput sequencing have been accomplished using different protocols (Kistler and Shapiro, 2011; Li *et al.*, 2011; Kistler, 2012; Fernández *et al.*, 2013). In our case, we developed a pipeline integrating the optimized PTB (*N*-phenacylthiazolium bromide) and guanidine hydrochloride DNA extraction from single wheat seeds and subsequent compatible whole-genome shotgun sequencing with Illumina HiSeq X10. Notably, besides reducing the probability of contamination from different seeds, our improved method with single-seed library construction and sequencing ensures no genomic heterogeneity caused by mixing different seeds, which cannot be used as a representative individual and should be avoided in population studies (Vohr *et al.*, 2015; Mascher *et al.*, 2016).

To demonstrate the efficacy of our current single-seed library construction, a DNA sample extracted from a mixture of the ancient seeds collected from all tombs was also extracted as a control. Subsequently, one mixture sample and seven samples each containing DNA extracted from a single seed were input following library construction and sequencing. In total, 17.1–260.3 million high-quality Illumina paired-end reads were generated (Table S3).

To estimate the content of endogenous DNA, the paired-end Illumina reads were aligned to the assembly of the reference genome of hexaploid common wheat (cv. Chinese Spring, genome BBAADD, IWGSC RefSeq v1.0; http://www.wheatgenome.org/News/Latest-news/RefSeq-v1.0-URGI). As expected, the mixed sample had higher percentages of total mapped and unique mapped reads against the reference genome (total mapping and unique mapping rates were 90.02% and 26.57%, respectively) than the single-seed samples (total mapping and unique mapping rates ranged from 5.38% to 89.36% and 9.12% to 20.99%, respectively), which were lower than both of the re-sequenced modern wheat samples (99.85% and 70%, Table S3). Notably, the M29-12-2 and M29-12-3 samples generated endogenous rates similar to the mixed sample in both total mapping rate and unique mapping rate. In comparison with the respective mapping rates in published ancient plant samples, the high total mapping and unique mapping rates of our single-seed samples indicated the sufficient endogenous content of DNA reads generated by current single-seed DNA extraction and sequencing methods (Kistler and Shapiro, 2011; Kistler, 2012; Fernández *et al.*, 2013).

According to the radiocarbon dating data, these ancient wheat seeds excavated at Xiaohe and Gumugou cemetery sites were estimated to be from approximately 3800–3200 years BP (Yang *et al.*, 2014; Zhang *et al.*, 2017). However, to demonstrate the authenticity of the ancient wheat DNA with no contamination from modern DNA, we examined the empirical sequence characteristics of these ancient DNA samples, which included short-length sequenced DNA fragments in ancient samples and abundant damage patterns (reflected as deamination-derived C-to-T and G-to-A nucleotide mismatches) at both the 5′- and 3′-ends of the sequenced DNA fragments. Specifically, most sequenced DNA fragments were short reads (44.37 ± 11.70), which were read through via current 150-bp PE Illumina reading (Figure S3). Additionally, all samples displayed enriched C-to-T and G-to-A patterns at the 5′- and 3′-end, respectively (Figure S2). For example, within ancient sample M29-19-2, we observed 15.02% and 14.97% C-to-T and G-to-A misincorporation at the 5′- and 3′-end, respectively (Figure S2), with a mean DNA fragment length of 44.98 bp (Figure S3). These characteristics of ancient samples were not detected in the re-sequencing data of modern wheat sample (Figures S2 and S3). These results strongly demonstrated the ancient origins of the DNA fragments extracted from each wheat seed, which were free of potential modern DNA contamination.

Sequenced reads covering all three subgenomes of common wheat confirmed the hexaploid identity of all ancient wheat samples

Considering that multiple factors can cause morphological variations in grain seeds, the foregoing morphological observations implicating the BBAADD genomic composition were still not conclusive. To ascertain the genomic composition of the ancient wheat seeds, we further dissected the sequencing reads of ancient and modern common wheat that uniquely mapped to the reference genome of hexaploid common wheat and summarized them in

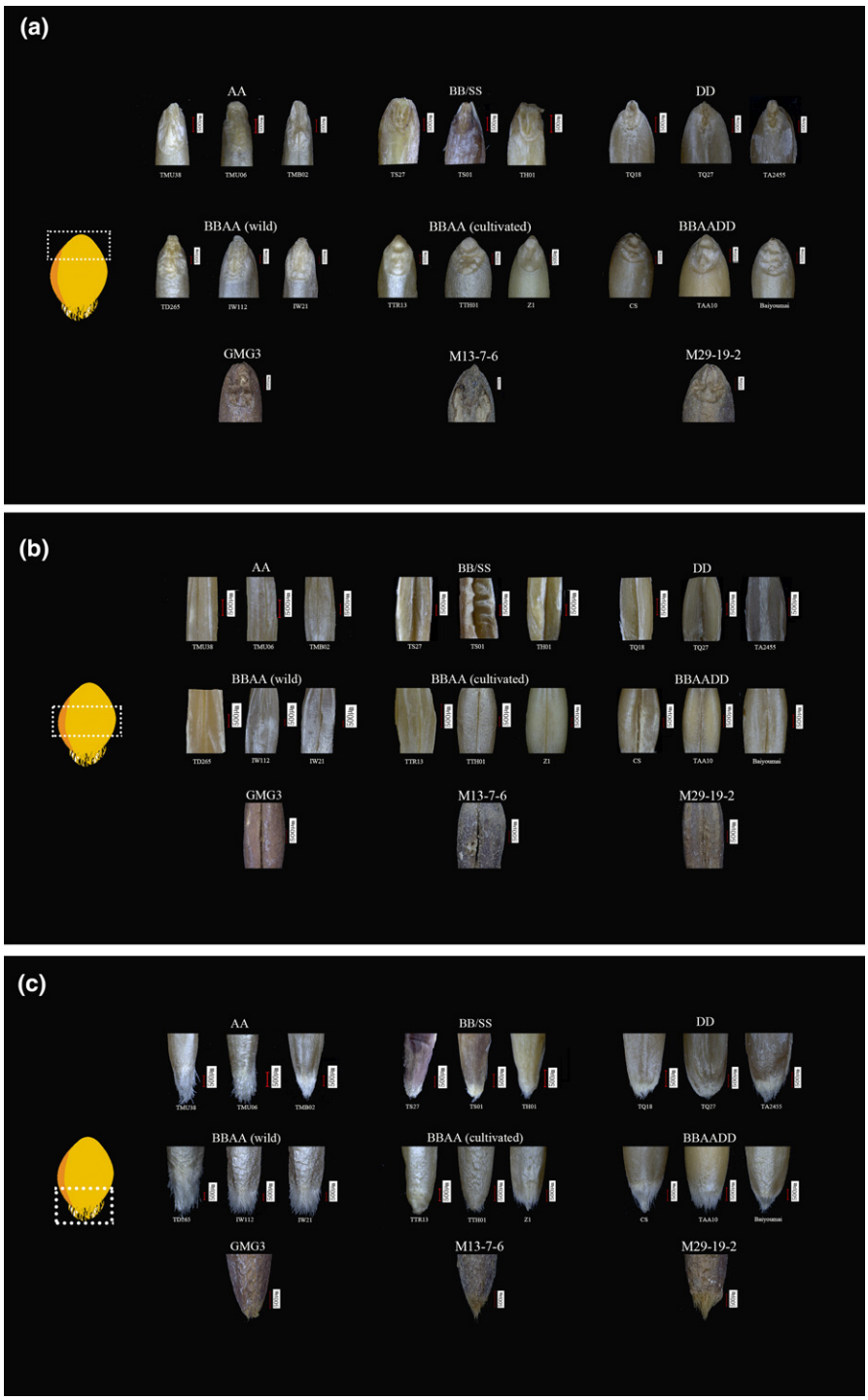


Figure 1. Microscopic observations of ancient and modern seeds. Ancient wheat seeds and other modern related samples were observed under a super-depth-of-field three-dimensional microscope. The characteristics of the upper embryo region (in panel a), around the ventral crease region (in panel b), and at the bottom tips (in panel c) are specifically magnified and compared among samples. Ancient wheat seeds include GMG3 (Gumugou cemetery), M13-7-6 (Xiaohe cemetery), and M29-19-2 (Xiaohe cemetery). Modern wheat and goatgrass species include AA genome wheat: *Triticum urartu* and *T. monococcum*; BB/SS genome goatgrass: *Aegilops speltoides* and *Ae. sharonensis*; DD genome goatgrass: *Ae. tauschii*; BBAA genome wild wheat: *T. dicoccoides*; BBAA genome cultivated wheat: *T. durum* and *T. carthlicum*; and BBAADD genome hexaploid common wheat: *T. aestivum*.

terms of their coverage across the three subgenomes (A, B, and D) of hexaploid common wheat. Consistent with modern common wheat, all three subgenomes in ancient wheat samples were unbiased in coverage by our sequenced ancient DNA reads (Figures 2 and S4). These results unambiguously confirmed that our samples were

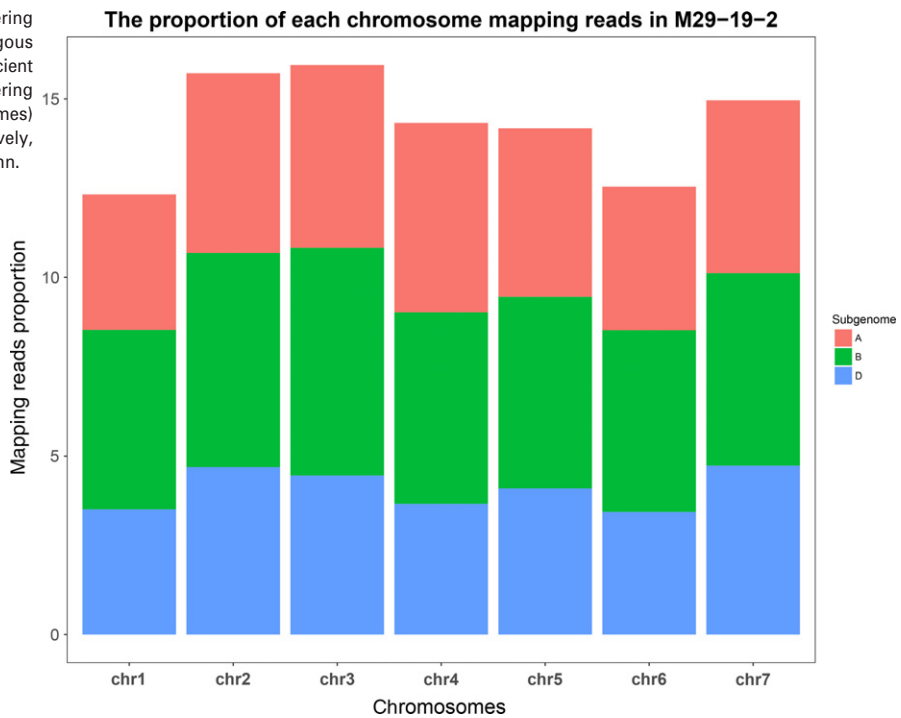


Figure 2. Proportions of mapped reads covering respective subgenomes within each homologous chromosome group. Within the M29-19-2 ancient sample, the proportion of mapped reads covering the three subgenomes (A, B, and D subgenomes) are coloured in red, green, and blue, respectively, and are stacked within each chromosome column.

hexaploid wheat rather than any diploid wheat progenitors or tetraploid wheat species. Moreover, the almost even distribution of mapped reads across all component chromosomes of each subgenome (Figure S7) allowed our further unbiased genomic ancestry analyses involving phylogenetic construction and population stratification (Table S2).

Well maintained ancient genomic ancestry within the common wheat landraces grown in the V-SWAS and IX-Q&T wheat-growing zones

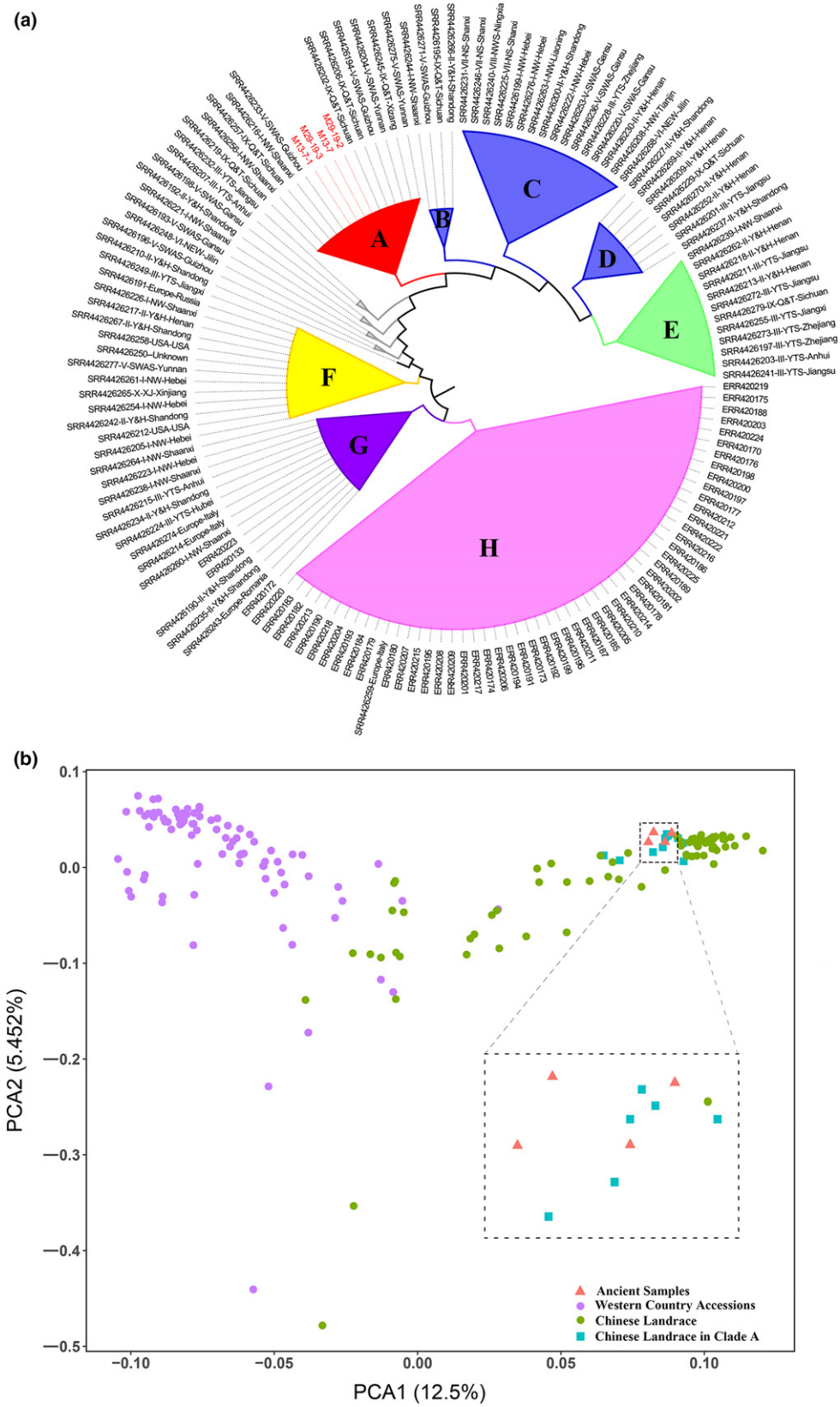
As noted above, common wheat was introduced to China approximately 4500 years ago, as supported by the hitherto oldest common wheat remnants uncovered at the 'Donghuishan' site in Gansu Province (dated to 4605 ± 150 and 4260 ± 80 BP) (Li and Mo, 2004; Flad *et al.*, 2010). Our ancient wheat samples with an estimated age of 3800–3200 BP are likely to constitute an ancient wheat population representing the status of ancient wheat after its initial introduction to China (Betts *et al.*, 2014; Yang *et al.*, 2014; Zhang *et al.*, 2017). Given that present-day common wheat is widely grown in China in 10 agroecological zones (Figure 3), we investigated whether the ancestry of the foregoing ancient wheat population had been maintained in any of the modern wheat landraces grown in China (He, 2001). Here, we included the present wheat accessions grown in Western countries (European and Australian regions) as the outgroup.

To address the question of ancestry, we initially measured the relatedness of the ancient samples with present-day common wheat landraces based upon their exonic nucleotide compositions. Public transcriptome data of representative modern common wheat landraces in Chinese agroecological zones and other Western countries were collected (Miller *et al.*, 2016; Wang *et al.*, 2017). By referring to the same common wheat genome assembly, the exonic SNPs identified in the respective landraces and accessions were concatenated into haplotypes and aligned. In total, the valid data set comprised 7058 SNPs (Table S2). A neighbour-joining phylogenetic tree was constructed based on the aligned haplotypes (Figure 4a). The phylogenetic clustering clearly divided the wheat accessions into eight collapsed clades (A–H clades). By considering the constituent landraces and accessions within each clade and their geographic origins, we further colour coded the clades and dissected their possible mutual relationships and degrees of relatedness with our ancient wheat samples (Figure 4a). Results showed that: (i) all ancient wheat samples resided in clade A coloured in red, which was mainly composed of the landraces grown in V-SWAS (Upper Yangtze valley, southwestern autumn-sown spring wheat zone, 4 landraces) and IX-Q&T (Qinghai-Tibetan plateau margins of highland China, spring–winter wheat zone, 5 landraces); (ii) clades B–D, denoted by dark blue, mainly contained the landraces locally grown in the I-NW (northern winter wheat zone, 6 landraces), II-Y&H (Yellow and Huai River valleys facultative wheat zone, 9 landraces), and V-SWAS (4 landraces); (iii) clade E, coloured in light green, was overrepresented by landraces grown in III-YTS (Middle and Lower Yangtze valley autumn-sown spring wheat zone, 7 landraces). All clades B–E were closely related to

Figure 3. Geographic locations of collected ancient wheat samples (Gumugou and Xiaohe cemeteries) and modern wheat landraces grown in 10 agroecological wheat zones in China. Different colours indicate the respective wheat-growing zones: I-NW (northern winter wheat zone), II-Y&H (Yellow and Huai River valleys facultative wheat zone), III-YTS (Middle and Lower Yangtze valley, autumn-sown spring wheat zone), IV-SAS (southern autumn-sown spring wheat zone), V-SWAS (Upper Yangtze valley, southwestern autumn-sown spring wheat zone), VI-NES (northeastern spring wheat zone), VII-NS (northern spring wheat zone), VIII-NWS (northwestern spring wheat zone), IX-Q&T (Qinghai-Tibetan plateau margins of highland China, spring–winter wheat zone) and X-XJ (Xinjiang winter-spring wheat zone). Sampling locations of modern landraces are indicated by asterisks which were supported by Wang *et al.*, 2017. There were no sequenced landrace samples available in IV-SAS and X-XJ.

clade A; (iv) clade F, shaded in yellow, was composed of a mixture of landraces grown in various Chinese zones and accessions from Western countries. It is speculated that the Chinese wheat landraces in clade F might harbour genomic sequences ascribed to gene flow/introgression from the accessions of Western countries; (v) clade G, coloured in purple, included Chinese landraces from I-NW, II-Y&H, and III-YTS and certain accessions from Western countries. Given the known historical periods of further wheat introductions from Western countries in modern times, the close clustering of clade G with the outgroup accessions from Western countries (clade H) implied that its component landraces could have been recently introduced from Western accessions; and (vi) as noted above, clade H, coloured with pink and mainly encompassing accessions of Western countries, was expected to be located distantly from clades A–F and served as the outgroup together with clade G (Figure 4a) (Dodson *et al.*, 2013; Betts *et al.*, 2014). Outgroup clustering of clade H also suggested automorphic evolution in Chinese wheat populations (including the ancient and present-day Chinese wheat samples) and/or in the Western wheat populations. The foregoing overall clade clustering was also supported by principal component analysis (PCA), in which the SNP divergence among all samples was further parsed into several principal components (Patterson *et al.*, 2007).

Figure 4. Ancient genomic ancestry in modern wheat accessions and landraces reflected by their phylogenic and clustering relatedness with the ancient samples. (a) The neighbour-joining phylogenetic tree was constructed based on effective exonic SNPs identified in all samples. Different clades are coloured in terms of the origins of their major component accessions and/or landraces: clade A, coloured in red, mainly includes the landraces grown in V-SWAS and IX-Q&T; clades B–D in dark blue mainly include the landraces locally grown in I-NW, II-Y&H and V-SWAS; clade E in light green is predominated by landraces grown in III-YTS; clade F, shaded in yellow, is composed of mixed landraces grown in various Chinese zones and accessions from Western countries; clade G, coloured in purple, includes Chinese landraces from I-NW, II-Y&H, and III-YTS and certain accessions from Western countries; and clade H in pink mainly encompasses Western accessions and serves as the outgroup together with clade G. (b) PCA clustering of ancient wheat samples and modern wheat accessions and landraces is plotted based on the same set of exonic SNPs. The inset magnifies the PCA space around the ancient samples that are closely clustered with Chinese landraces in phylogenic clade A.



As shown, the second principal component (PC2) clearly categorized ancient wheat samples and landraces grown in China into the same large group, which was clearly separated from the accessions from Western countries (Figure 4b). Collectively, at the haplotype level reflected by the exonic SNP composition, our ancient wheat samples displayed closer relatedness to the modern landraces, especially to those grown in southwestern China (V-SWAS and IX-Q&T).

In addition to above nonparametric approaches evaluating the phylogenetic relatedness and ancestry, we also performed model-based estimation of ancestry and evaluated the population structure within ancient wheat samples and present-day wheat samples using ADMIXTURE with a certain number of ancestral populations (*K*) (Alexander *et al.*, 2009). For each value of *K*, we estimated the cross-validation error based on the maximum likelihood solution across replicates. We observed minimum cross-validation error estimates when $K = 3$, and had the lowest value among other trials of different *K* values (Figure S5). We present admixture proportions for $K = 3$ (minimum cross-validation error). This population stratification also clustered our ancient and modern wheat samples into their origin wheat zones (Figures 5 and S6). We found the following: (i) the structure of genetic diversity in modern Chinese landraces grown in V-SWAS and IX-Q&T was to a large extent consistent with that in the ancient wheat population; (ii) in addition to the ancient genetic structures (in red), landraces in I-NW, II-Y&H, and III-YTS had their own featured modern structures shown in mosaic red and green; and (iii) accessions from Western countries displayed unique structures shown in blue. All these observations also support the hierarchical ancestry maintained in different modern Chinese wheat landraces.

Model-based allelic composition corroborates the archaic spreading of common wheat to the Qinghai-Tibetan plateau margins and subsequent entry into the Yangtze River valley

Previous dating records of archeological wheat remains implicated the eastward spreading of common wheat into central China (around ca. 4500 BP, I-NW and II-Y&H, route 1 in Figure 6) and later spread into Qinghai-Tibetan plateau margins of highland China (around ca. 3400–3300 BP, IX-Q&T, route 2 in Figure 6), both of which initiated from the Gansu center (around ca. 5000 BP, Figure 6) (Betts *et al.*, 2014). Available whole-genomic sequencing data of both ancient and modern wheat populations could help us verify those spreading routes as well as offer new insights about further dispersal across the wheat zones without any excavated archeological samples. Accordingly, we conducted a model-based evaluation of the population structure within ancient wheat samples and present-day wheat landraces using D statistical as well as f_3 analyses (Patterson *et al.*, 2012). All possible proxies were inserted in the D-statistics in the form D (P1, P2; P3, O), in which modern landraces were assigned to P1 and P2, whereas the P3 and O were fixed as the ancient wheat and Western wheat accessions, respectively. Similarly, within each f_3 statistical analysis [f_3 (X, Ancient; Western Countries), Figure 7], modern Chinese landraces were assigned at the X position.

To infer the spreading routes based on the results of f_3 and D statistical analyses, our underlying assumption/hypothesis was the well preserved hierarchical allelic composition/frequency within the dispersal of common wheat population since ancient times. As noted and summarized, modern landraces grown in different wheat zones exhibited hierarchical similarity in allelic frequency with the

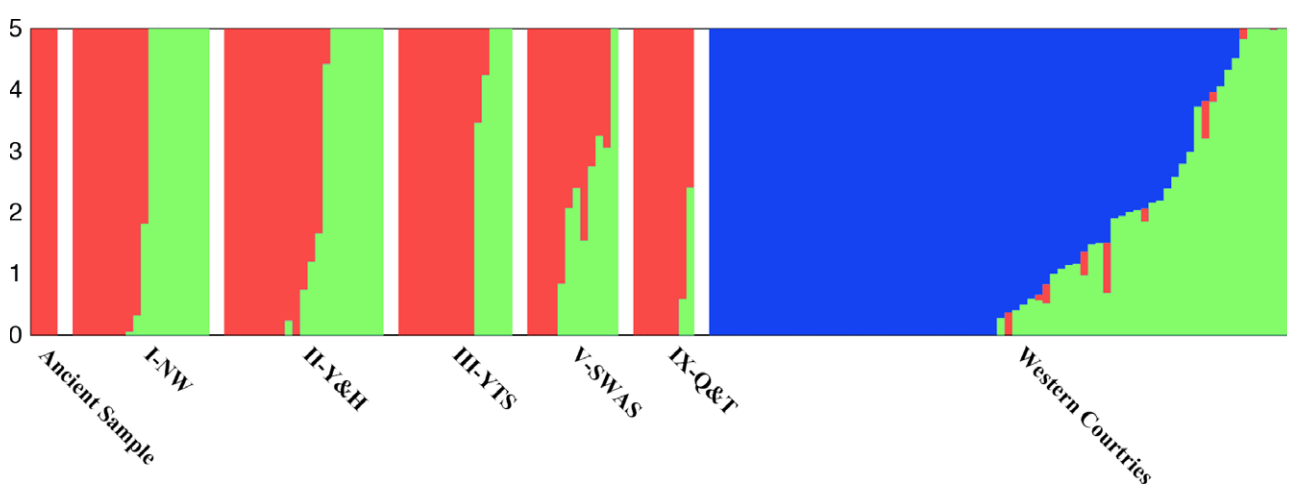


Figure 5. Population structure of ancient wheat samples and modern common wheat accessions and landraces based on modelled ADMIXTURE analysis at $K = 3$. The component topological structural groups of modern wheat are divided according to wheat-growing zones. Wheat-growing zones with more than three representative modern landraces are included in this ADMIXTURE analysis.

Figure 6. Possible routes for the entry and spread of wheat in China Two main entry routes of wheat in China are represented with red and blue dotted arrowhead lines, separately. Three postulated spread routes of wheat are denoted in shaded arrows: eastward spread route 1 involves wheat spread into central China (I-NW and II-Y&H); southwestern spread route 2 involves dispersal into Qinghai-Tibetan plateau margins (IX-Q&T); spread route 3, postulated in our study, involves wheat spread into Yangtze River valley (III-YTS and V-SWAS). Listed ages of all entry and spread routes were inferred based on previous studies (Betts *et al.*, 2014).

ancient wheat, which was ordered as follows: IX-Q&T > V-SWAS ≈ III-YTS > II-Y&H ≈ I-NW (Figure 7 and Table S4). The greatest similarities in allelic frequency between modern landraces of IX-Q&T and ancient wheat corroborate foregoing archaic spreading of common wheat into Qinghai-Tibetan plateau margins of highland China (Figure 7 and route 2 in Figure 6), which also confirms the well maintained ancestry subsequent to their introduction into China. Intriguingly, although the eastward dispersal occurred longer ago (around ca. 4500 BP, route 1 in Figure 6), the least similarity in allelic frequencies of modern landraces in central China relative to the ancient wheat (Figure 7) and their similar population structures infer a minimum ancestry maintained in the admixture of common wheat in central China. Furthermore, consistent less similar allelic frequencies in III-YTS and V-SWAS with that in ancient wheat implicates the subsequent archaic entry into Yangtze River valley (III-YTS and V-SWAS, route 3 in Figure 6).

DISCUSSION

As one of the most important food crops in the world, common wheat was initially domesticated in the Fertile Crescent approximately 10 000 years ago and subsequently spread westward to Europe and eastward to China approximately 4500 years ago. Questions about the dispersion history of ancient wheat within China after introduction, the extent to which the ancient genomic ancestry still maintained in present-day local Chinese wheat landraces, and the local adaptive evolution and selection of Chinese wheat landraces are still underexplored at the population level. Given that genetic analyses of ancient DNA can greatly inform research into these issues, our study employed well preserved ancient wheat samples to develop an efficient whole-genomic shotgun sequencing pipeline, which is suitable for analyzing each excavated archeological wheat seed. Based upon this technical platform, we further dissected the genomic ancestry of ancient wheat retained in modern representative Chinese wheat landraces, estimated their extent of similarity in population structure and allelic composition with ancient wheat, evaluated the potential population mixture among related wheat populations, and inferred the potential spreading route of common wheat in China.

Considering the morphological changes resulting from ancient preservation, molecular methods have been widely adopted to ascertain the genomic identity of archeological plant samples of interest in population studies (Li *et al.*, 2015; Bilgic *et al.*, 2016; Fornaciari *et al.*, 2016; Mascher *et al.*, 2016; Wales *et al.*, 2016; Di *et al.*, 2018). Our Xiaohe ancient wheat seeds were previously determined to be hexaploid based on their longer amplified D-subgenomic rDNA IGS (intergenic spacer region), which should be absent from the respective regions in the A and B subgenomes (Li *et al.*, 2011). Without addressing the composition of the B and A subgenome, the conventional amplification of the D-subgenomic region could not exclude the possibility of its being a D-genome diploid *Aegilops tauschii* also harbouring the D-genome. Therefore, our whole-genomic shotgun sequencing revealed

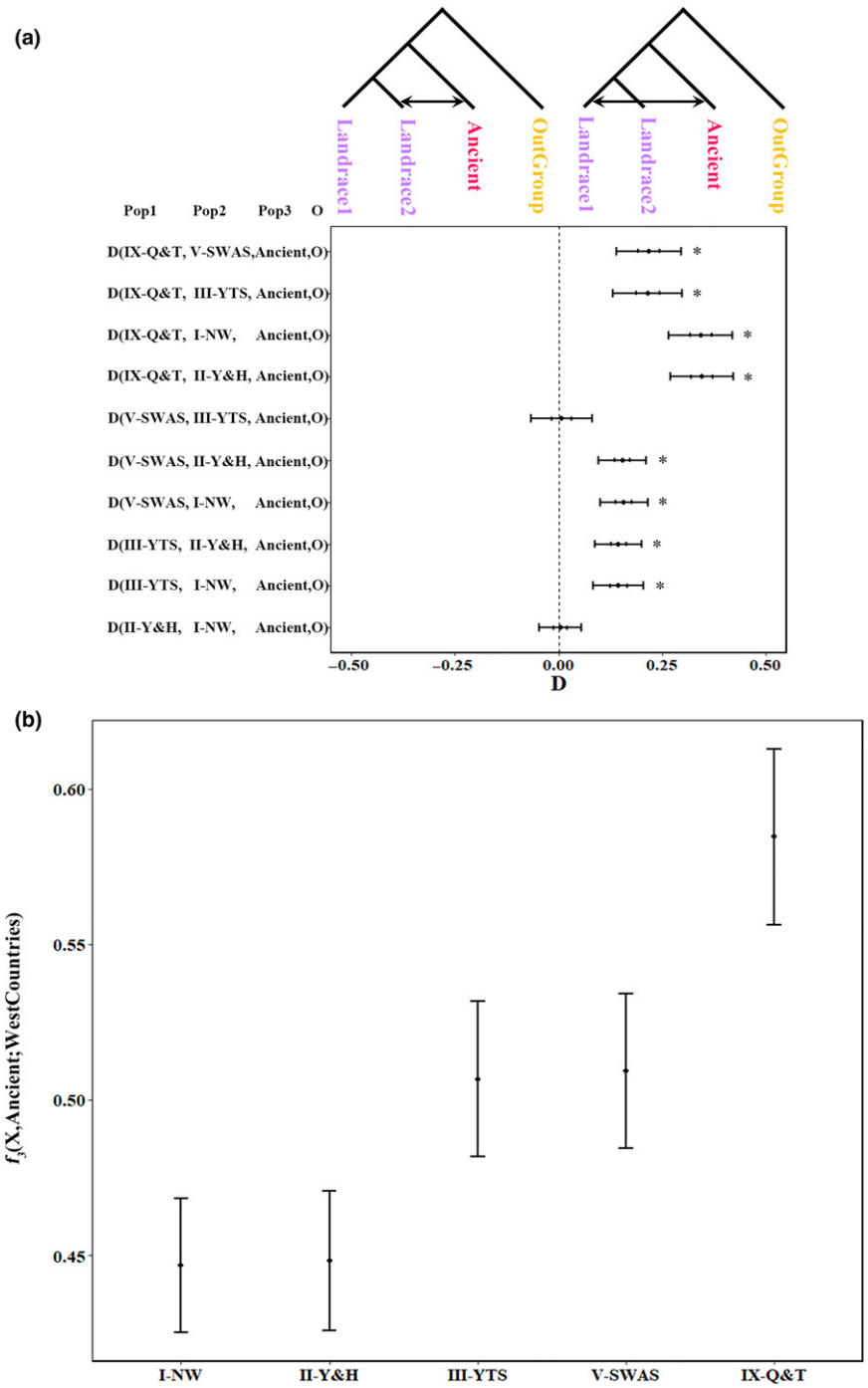


Figure 7. D-statistics and f_3-statistics for different wheat populations. (a) Ancient and modern common wheat populations are classified into a quadruple, accordingly. Western country accessions (in clade H) are selected as the outgroup and introgression accessions (in clade F) and introduction accessions (in clade G) are excluded. Positive D values indicate that P1 shares more derived alleles with P3 than does P2, whereas negative statistics indicate that P2 shares more alleles with P3 than does P1. Black error bars correspond to 1 and 3 standard errors. All significant results are denoted with asterisks (corresponding $|Z| > 3$). (b) F_3-statistics for Chinese landraces which show the extent of shared drift with ancient samples compared with Western country accessions. Population X (modern Chinese landraces grown in the specific agroecological zone) with higher f_3 values than zero indicates that population X is closer related to ancient samples relative to those Western country accessions. Bars dissecting the points show standard error.

ancient DNA reads derived from all three component subgenomes, providing unambiguous evidence confirming that our ancient wheat is hexaploid common wheat.

As an important crop introduced into China, common wheat landraces have been adapted to diverse local environments (Dwivedi *et al.*, 2016). In addition to being a

model polyploid species to characterize genomic evolution after genome doubling, increasing attention has been given to the evaluation of the genetic diversity of present Chinese wheat landraces, the characterization of their population structure, and the identification of potential genomic loci under selection in each local growing zone. However, there was no evidence for whether and to what extent these present Chinese landraces could still maintain the genomic ancestry of ancient wheat populations after introduction. Therefore, our inference of genomic ancestry based on both phylogenetic relatedness and PCA categorization initially characterized the local landraces with potential recent introduction and gene flow/introgression. More importantly, together with the characterized population structures, the results corroborated the well maintained ancient genomic ancestry within the modern common wheat, especially in those grown in southwestern China (V-SWAS and IX-Q&T). Our observed well maintained genomic ancestry indicates that the present landraces in southwest should harbour higher diversity than those grown in other agroecological zones (Appels and Lagudah, 1990; Wang *et al.*, 2013; Zhou *et al.*, 2018). As revealed by the exonic SNPs composition, the foregoing expectation is supported by the highest heterozygosity in landraces of V-SWAS relative to those in other wheat-growing zones (Zhou *et al.*, 2018).

Previously, it has been established that natural adaptation to different environmental factors in different growth zones and human selection for improved agronomical traits could interact and contribute to the genetic variation of wheat after introduction into China (Huw *et al.*, 2008; Dotlačil *et al.*, 2010; Dwivedi *et al.*, 2016; Zhou *et al.*, 2018). Therefore, the current ancestral relationship could be underpinned by these two integrated factors. Specifically, for natural adaptation, it was documented that during the period 9700–3700 BP, the Luobupo region surrounding the Xiaohe and Gumugou zones had warm and wet conditions similar to the current climates in southwestern regions (Xia and Zhao, 2005). The ancestral genomic composition maintained in wheat of the southwestern regions could still allow their adaptation to natural habitats similar to their ancient progenitors; selection by similar natural habitats could have shaped their similar population structures as well (Belay *et al.*, 1995; d'Alpoim Guedes *et al.*, 2015). With human selection, the historical development of agriculture in central China (corresponding to I-NW and II-Y&H) may have greatly diluted and blurred the genomic ancestry of their local wheat landraces via hybridization, introgression, and even introduction of wheat cultivars from other origins (Zhuang, 2003; He *et al.*, 2011). This inference was also supported by the outlier landraces clustered in clades F and G (Figure 4a). Conversely, the less intensified hybridization breeding in wheat in southwestern regions might have facilitated conservation of the ancient genomic ancestry.

The exact routes for the spreading of wheat within China are still contentious (Dodson *et al.*, 2013; Betts *et al.*, 2014; Weiss *et al.*, 2015). To address this issue, several possible scenarios were postulated, which were based upon the dating records of archeological wheat remains or upon analyzing the genetic structure of only modern Chinese landraces (Dodson *et al.*, 2013; Betts *et al.*, 2014). As noted, neither of those two methods are absolutely reliable. Specifically, the accuracy of former archeological dating relies on the quality of the remains and the availability of ancient remains (Dodson *et al.*, 2013; Betts *et al.*, 2014). Accordingly, more excavated available remains could reach different spreading routes. For the latter method, valid gene flow inferences need appropriate reference populations (Pickrell and Pritchard, 2012; Da Fonseca *et al.*, 2015). Together with those Western accessions as outgroup population, our ancient wheat seeds offer an ideal reference population to trace the direction and sites involved in the wheat spreading route in China. Although not all putative node sites in the postulated spreading routes were involved in this study (wheat-growing zones with less than three landrace samples were excluded, Experimental procedures), previous archaic spreading of common wheat into Qinghai-Tibetan plateau margins of highland China (around ca. 3400–3300 BP, route 2 in Figure 6; Betts *et al.*, 2014) is fully supported by our molecular genetic evidence at the population level. This spreading route is also consistent with the previous finding showing that common wheat could have become an important subsistence for the local people in eastern margins of the Tibetan plateau shortly after its introduction (Guedes *et al.*, 2015). However, as for the reported route 1 (Figure 6; Betts *et al.*, 2014), in which the ancient eastward spreading of wheat (around ca. 4500 BP) into I-NW and II-Y&H of central China, it seems contradictory with their least retained ancestry and allelic similarity inherited from our ancient samples (around 3800–3200 BP). Two possible scenarios could explain this: (i) it is possible that the wheat involved in these two dispersal events (eastward spread into central China and later spread into Qinghai-Tibetan plateau margins of highland China) could be genetically distinct at the population level. Accordingly, the local landraces grown in central China could still maintain the ancestry and population structural features from the ancient dispersal event (around ca. 4500 BP); and (ii) as mentioned previously, advanced agriculture in central China may have greatly altered the allelic composition via hybridization, introgression, and even external introduction of wheat cultivars (Zhuang, 2003; He *et al.*, 2011).

Besides the routes 1 and 2 proposed previously, based upon the data generated in our study, another dispersal route of common wheat from Qinghai-Tibetan plateau into

the Yangtze River valley (III-YTS and V-SWAS, route 3 in Figure 6) was proposed here, which implicates the crucial role of those pioneering populations in establishing common wheat as staple across the vast regions of China. In addition, although there are no archeological remains excavated nearby and/or along the Yangtze River valley, our study still offers important insights about the origin of the wheat grown in the respective wheat zones. Of course, it is still necessary to include more ancient wheat seeds of different ages excavated in central China to perform a more comprehensive population analysis, such as the Treemix T_F evaluation, to confirm which scenario is the likeliest and finalize the exact dispersal route(s) of common wheat in China.

EXPERIMENTAL PROCEDURES

Archeological context

Desiccated wheat seeds were excavated from Xiaohe and Gumugou cemeteries, which are located in the Taklimakan Desert in the Xinjiang Uyghur Autonomous Region of northwestern China (Figure 3). Previous archaeological researches have shown that wheat and barley were cultivated in Xiaohe and Gumugou regions (Yang *et al.*, 2014, Betts *et al.*, 2019). The wheat seeds used in this study were obtained from three tombs in Xiaohe cemetery (designated as M13, M17, and M29) and two tombs in Gumugou cemetery (designated as M1 and M3). The ancient samples excavated from Xiaohe and Gumugou cemeteries were ^{14}C radiocarbon dated with a direct accelerator mass spectrometer (AMS) to approximately 3800–3200 BP (Yang *et al.*, 2014; Zhang *et al.*, 2017).

Morphological observation

Detailed morphological observations were performed to roughly discriminate the samples such as wheat seeds and discriminate them specifically in terms of diploid, tetraploid and hexaploid wheat, which have different morphological characteristics. Ancient samples together with modern wheat seeds were observed and photographed with a super-depth-of-field three-dimensional microscope (×30 magnification). Detailed and careful observations were performed for all samples at specific subdivisions, including the upper embryo region, middle endosperm crease region, and the bottom tip region.

DNA extraction, library construction and sequencing

Due to the exceptionally arid conditions, most of the wheat remains were well preserved and considered to be suitable for DNA-based studies. Therefore, seven samples, each composed of a single wheat seed in dry plump and oval shape, were utilized for DNA extraction and subsequent library preparations for shotgun sequencing. In addition, a mixed sample consisting of 12 wheat seeds as the control was included to test the efficacy of DNA extraction and library construction from single seeds.

Within our method of DNA extraction, appropriate precautions were taken to ensure the authenticity of the ancient DNA results (Li *et al.*, 2011). Briefly, DNA extraction steps (including grinding of grain, DNA extraction and polymerase chain reaction (PCR) set-up) were carried out in separate rooms entirely dedicated to ancient DNA studies. All consumables were cleaned with DNA-OFF and ultraviolet (UV) light irradiated before use. For the ancient DNA extraction, each ancient sample was soaked in 10% sodium hypochlorite solution for 10 min, rinsed with absolute ethanol, and allowed to air dry under UV-irradiation in clean-room facilities. After grinding into powder in liquid nitrogen, 5–10 mg of powder from each sample was collected in 1.5 ml tube for the DNA extraction and 1 ml extraction buffer containing 1% SDS together with 10 mM ethylene diamine tetraacetic acid (EDTA) pH 8.0, 10 mM Tris-Cl pH 8.0, 15 mg/ml proteinase K, N-PTB and 5 mM NaCl was added. After suspension by vortexing and overnight incubation at 37°C, the extracts were centrifuged for 2 min at 12500 ***g*** and the supernatant was transferred into a 50 ml tube equipped with a binding apparatus reservoir containing 13 ml binding buffer (including guanidine hydrochloride, isopropanol, sodium acetate and Tween-20). After centrifuging for 7 min at 170 ***g***, the remaining supernatant was loaded into a MinElute column (Qiagen, Valencia, CA, USA, ID: 28006) and 750 μl PE buffer was added to the collection tube. After briefly centrifuging, the flow-through was discarded. This PE washing step was repeated, and the MinElute column was dried for 1 min. Next, 25 μl TET (Tris−Cl, EDTA and Tween-20) was added to the column, which was then centrifuged at maximum speed for 30 sec. After repetition of this collection step, approximately 60 μl extracted DNA was collected for further study.

For the shotgun library preparation and sequencing, libraries compatible with Illumina HiSeq sequencing were built following a well established protocol from Meyer and Kircher (Kircher *et al.*, 2012). Some specific modifications were made, which included: (i) 20 μl aliquot of each DNA extract was used for constructing Illumina libraries; (ii) library amplification was performed using Q5® High-Fidelity DNA Polymerase; (iii) amplification conditions were 30 sec at 98°C; 14 cycles of 10 sec at 98°C, 75 sec at 60°C with a final extension at 60°C for 6 min, and holding at 4°C; and (iv) amplified libraries were cleaned using Agencourt® AMPure XP beads with a ratio ranging from 1:1.5 to 1:1.8 (library volume:bead volume) and were finally eluted in 26 μl sterile water. These libraries were subsequently used for shotgun sequencing. Multiplex shotgun sequencing was carried out using an Illumina HiSeq X10 platform at Novo Inc., Beijing, in PE 150 mode.

DNA damage patterns and SNP calling of high-quality mapped sequencing reads

As for the ancient DNA samples, the adapter sequences of raw Illumina sequencing reads were removed, and overlapping paired-end reads were merged using Adapter Removal 2.2.2 with the shortest overlap at 11 bp (Schubert *et al.*, 2016). The merged reads shorter than 30 bp or average quality below 20 were discarded. All merged high-quality reads were aligned to assembly Chinese Spring (IWGSC RefSeq v1.0) with BWA-MEM 0.7.7 in default parameters (Li, 2013). SAM format mapping files were converted to BAM format and then sorted using SAMtools 0.1.17 (Li *et al.*, 2009). The duplicated reads were removed by employing dedup (Peltzer *et al.*, 2016). As a relative control, the re-sequencing reads of modern common wheat (Accession Name: Renan; SRA: ERR1706944) were also downloaded, processed, and mapped by the same method as described above. DNA damage patterns of all ancient and modern common wheat DNA samples were assessed by examining the C-to-T substitution rates toward the sequencing 3′-end and complementary G-to-A rates toward the 5′-end using MapDamage 2.0 (Jónsson *et al.*, 2013). SNP calling was completed by SAMtools mpileup (with SguD parameters) and bcftools view (with cvNg parameters) for reads with mapping quality greater than 30 (*Q* value >30) and nucleotide sequencing quality above 20 (Li, 2011). Based on the composition of identified

SNPs, the reads covering the three subgenomes (A, B, and D subgenomes) were summarized separately.

Ancient genomic ancestral analyses for the modern wheat grown in China

To trace the ancient genomic ancestry in extant wheat grown in China, transcriptomic data of other common wheat were downloaded from NCBI (Accession Nos.: SRP091625 and ERP004689), which included 74 landraces and 16 elite varieties and 100 common wheat accessions grown locally in China and European countries. All the geographic information of landraces in China were detailed in Wang *et al.*, 2017. The transcriptomic reads of all the extant accessions of wheat were also aligned to the same reference with BWA-MEM 0.7.7, and their SNP calling was completed as described above for ancient DNA sequencing data.

SNPs identified in ancient and extant wheat samples were further filtered to ensure the inclusion of sufficient informative SNP positions covering the genomes of all samples of interest as follows: (i) variants must be biallelic SNPs, (ii) at least 70% of extant modern wheat samples must have respective reads covering the corresponding SNP position; and (iii) one of the ancient samples must have read coverage.

Informative qualified SNPs were positionally concatenated and aligned, and the results were used to construct a neighbour-joining tree with the Jukes-Cantor distance model in MEGA 6.0 (Tamura *et al.*, 2013). Bootstrap values based on 1000 random samples were estimated for each branch node. In addition, PCA was performed with EIGENSOFT 6.1.4 for the qualified SNPs identified in ancient wheat samples and extant modern wheat accessions (Patterson *et al.*, 2007).

In addition to the previous nonparametric methods, model-based clustering analysis using ADMIXTURE was performed to characterize the population structure within ancient and modern wheat populations (Alexander *et al.*, 2009). Those wheat zones with more than three representative modern landraces were included in the ADMIXTURE analysis. Here, we used PLINK to remove SNPs in strong linkage disequilibrium, employing a window of 200 SNPs advanced by 25 SNPs and with an r^2 threshold of 0.4 (Purcell *et al.*, 2007). To identify the optimal K value with minimum cross-validation error, we ran ADMIXTURE with the cross-validation (–cv.) flag specifying from $K = 1$ to $K = 20$ clusters (with 100 replicates for each value of K). For each K value, the replicate with the highest log likelihood was retained.

To explore the topology between modern China wheat, ancient wheat and west wheat, we performed f_3-statistic (P1, P2; P3) using 3PopTest programs from the AdmixTools package. The correct topology defines that the value $f_3 > 0$ and the Z-score was of magnitude greater than 3. A most profitable model f_3 (X, Ancient; west countries) was adopted in our study, which generated all significant Z-score ($Z > 20$).

To compare the allelic composition among wheat individuals and populations, D-statistics with designed quadruples were calculated using AdmixTools based on the SNP matrix converted to the EIGENSOFT format (Patterson *et al.*, 2007). Being similar as above, those wheat zones with more than three representative modern landraces were included in the D-statistic analyses. Within each D-statistic quadruple modelled as D (P1, P2; P3, O), O represented the outgroup population which was set as the Western wheat accessions; P3 was assigned to the ancient wheat; P1 and P2 involved modern wheat populations in certain combination. Positive D values indicate that P1 shares more derived alleles with P3 than P2 does, while negative statistics indicate that P2 shares more alleles with P3 than P1 does. D-statistics with respective $|Z| > 3$ were accepted to be statistically significant.

DATA STATEMENT

All related sequencing data have been submitted to SRA database in NCBI under accession number SRP188209. Transcriptomic data of other common wheat were downloaded from NCBI (SRA: SRP091625 and ERP004689). Resequencing data of modern common wheat were downloaded from NCBI (SRA: ERR1706944).

ACKNOWLEDGEMENTS

We are grateful to the Xinjiang Cultural Relics and Archaeology Institute for providing us with the ancient wheat materials. This work was supported by grants from the National Key Research and Development Programme of China (2016YFD0101004), and the National Natural Science Foundation of China (grant no. 41472024), the Recruitment Programme of Global Youth Experts, the Programme of Changbai Mountain Scholars, and the Fundamental Research Fund for the Central Universities (2412017BJ005). We appreciate the knowledge and trainings given by the course of Evolution Biology in Northeast Normal University. We are grateful to critical comments by two anonymous reviewers for improving the manuscript.

AUTHOR CONTRIBUTIONS

Xiyan Wu, Baoxu Ding, Yinqiu Cui and Lei Gong planned and designed the research. Wenying Li and Yinqiu Cui collected the sample. Xiyan Wu, Chao Ning and Fan Zhang performed experiments. Baoxu Ding, Xiyan Wu, Bingqi Zhang, Jiaojiao Feng, Yibing Wang, Zhang Qun and Quanchao Zhang performed morphological observation. Baoxu Ding, Bingqi Zhang, Jiaojiao Feng, Yibing Wang, Haidan Wu, Ning Li, Zhibin Zhang, Xuhan Sun performed bioinformatic and genetic analysis. Xiyan Wu, Baoxu Ding, Bao Liu, Yinqiu Cui and Lei Gong wrote the manuscript. Xiyan Wu and Baoxu Ding contributed equally as first authors. Yinqiu Cui and Lei Gong contributed equally as corresponding authors.

CONFLICT OF INTEREST

The authors declare no competing interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Morphological features (length width and thickness of seed) in the ancient and modern wheat seeds.

Table S2. The proportion of each chromosome mapping reads in ancient samples.

Table S3. Summary of ancient wheat samples sequenced by Illumina HiSeq X10. Modern wheat Renan was selected as a control.

Table S4. F3/D statics for different comparisons between ancient wheat and modern wheat from Chinese Landrances and Europe. Standard errors and Z scores are given in columns SE and Z.

Figure S1. Photographs of seven ancient samples. The figures represented the front (a) and the ventral aspects (b) of wheat. Ancient wheat seeds include GMG1 and GMG3 (Gumugou cemetery),

M13-7-1 and M13-7-6 (Xiaohe cemetery), M17-12-7 (Xiaohe cemetery) and M29-19-2 and M29-19-3(Xiaohe cemetery).

Figure S2. Nucleotide misincorporation profiles in reads mapped to the whole-genome assembly of modern hexaploid common wheat. For each ancient sample, the C-to-T and G-to-A misincorporation frequencies (on the *y*-axis) at each nucleotide position (on the *x*-axis) at the 5′- (left panel) and 3′-end (right panel) are summarized and illustrated with red and blue lines, respectively. The modern wheat Renan is selected as a control.

Figure S3. The fragment length distribution of ancient samples. For each ancient sample, the number of fragments (on the *y*-axis) at each length category (on the *x*-axis) is presented. The modern wheat Renan is selected as a control.

Figure S4. Proportions of mapped reads covering respective subgenomes within each homologous chromosome group. Within each ancient sample, the proportion of mapped reads covering three subgenomes (A, B, and D subgenomes) is coloured in red, green, and blue, respectively, which are stacked within each chromosome column. The modern wheat Renan is selected as a control.

Figure S5. Cross-validation errors of the ADMIXTURE analysis. For each *K* value, we estimated the cross-validation error based on the maximum likelihood solution across replicates. The minimum cross-validation error estimate is reached when $K = 3$.

Figure S6. Population structure of ancient wheat samples and modern common wheat accessions and landraces based on modelled ADMIXTURE analysis from $K = 2$ to $K = 7$ (excluding $K = 3$). The component topological structural groups of modern wheat are divided according to wheat-growing zones. Wheat-growing zones with more than three representative modern landraces are included in this ADMIXTURE analysis.

Figure S7. The average depth along each chromosome. With each ancient sample, the average read depth in each 1Mbp window sliding along each chromosome was calculated and illustrated in grey dots. The fitted regression lines are represented (in blue). The modern wheat Renan was selected as a control.

REFERENCES

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664.

d'Alpoim Guedes, J.A., Lu, H., Hein, A.M. and Schmidt, A.H. (2015) Early evidence for the use of wheat and barley as staple crops on the margins of the Tibetan Plateau. *Proc. Natl Acad. Sci. USA*, **112**, 5625–5630.

Appels, R. and Lagudah, E.S. (1990) Manipulation of chromosomal segments from wild wheat for the improvement of bread wheat. *Funct. Plant Biol.* **17**, 253–266.

Belay, G., Tesemma, T., Bechere, E. and Mitiku, D. (1995) Natural and human selection for purple-grain tetraploid wheats in the Ethiopian highlands. *Genet. Resour. Crop Evol.* **42**, 387–391.

Betts, A., Jia, P.W. and Dodson, J. (2014) The origins of wheat in China and potential pathways for its introduction: a review. *Quatern. Int.* **348**, 158–168.

Betts, A., Jia, P. and Abuduresule, I. (2019) A new hypothesis for early Bronze Age cultural diversity in Xinjiang, China. *Archaeol Res Asia*, **17**, 204–213.

Bilgic, H., Hakki, E.E., Pandey, A., Khan, M.K. and Akkaya, M.S. (2016) Ancient DNA from 8400 Year-Old Catalhoyuk wheat: implications for the origin of neolithic agriculture. *PLoS ONE*, **11**, e0151974.

Cavanagh, C.R., Chao, S., Wang, S. *et al.* (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl Acad. Sci. USA*, **110**, 8057–8062.

Da Fonseca, R.R., Smith, B.D., Wales, N., Cappellini, E., Skoglund, P., Fumagalli, M., Samaniego, J.A., Carøe, C., Ávila-Arcos, M.C. and Hufnagel, D.E. (2015) The origin and evolution of maize in the Southwestern United States. *Nat. Plants*, **1**, 14003.

Di, A.D., Filippone, E., Ercolano, M.R. and Frusciante, L. (2018) Genome sequencing of ancient plant remains: findings, uses and potential applications for the study and improvement of modern crops. *Front. Plant Sci.* 9.

Dodson, J.R., Li, X., Zhou, X., Zhao, K., Sun, N. and Atahan, P. (2013) Origin and spread of wheat in China. *Quatern. Sci. Rev.* **72**, 108–111.

Dotlačil, L., Hermuth, J., Stehno, Z., DvořÁček, V., Bradová, J., Leišová, L. and Holubec, V. (2010) How can wheat landraces contribute to present breeding? *Czech J. Genet. Plant Breed.* **46**, S70–S74.

Dvorák. (1976) The relationship between the genome of Triticum urartu and the A and B genomes of Triticum aestivum. *Can. J. Genet. Cytol.* **18**(2), 371–377.

Dvorák, 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.

Dvorák, J., Akhunov, E.D., Akhunov, A.R., Deal, K.R. and Luo, M.C. (2006) Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Mol. Biol. Evol.* **23**, 1386–1396.

Dwivedi, S.L., Ceccarelli, S., Blair, M.W., Upadhyaya, H.D., Are, A.K. and Ortiz, R. (2016) Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Sci.* **21**, 31–42.

Feldman, M. (1976) Wheats: Triticum spp. (Gramineae-Triticinae). In *Evolution of Crop Plants* (Simmonds N.W., ed). London: Longmans, pp. 120–128.

Fernández, E., Thaw, S., Brown, T.A., Arroyo-Pardo, E., Buxó, R., Serret, M.D. and Araus, J.L. (2013) DNA analysis in charred grains of naked wheat from several archaeological sites in Spain. *J. Archaeol. Sci.* **40**, 659–670.

Flad, R., Li, S.C., Wu, X.H. and Zhao, Z.J. (2010) Early wheat in China: results from new studies at Donghuishan in the Hexi Corridor. *Holocene*, **20**, 955–965.

Fornaciari, R., Fornaciari, S., Francia, E., Mercuri, A.M. and Arru, L. (2016) Panicum spikelets from the Early Holocene Takarkori rockshelter (SW Libya): archaeo-molecular and -botanical investigations. *Plant Biosyst.* **152**, 1–13.

Guedes, J.A.D.A., Lu, H., Hein, A.M. and Schmidt, A.H. (2015) Early evidence for the use of wheat and barley as staple crops on the margins of the Tibetan Plateau. *Proc. Natl Acad. Sci. USA*, **112**, 5625–5630.

Gugerli, F., Parducci, L. and Petit, R.J. (2005) Ancient plant DNA: review and prospects. *New Phytol.* **166**, 409–418.

He, Z. (2001) *A History of Wheat Breeding in China*. Mexico, D.F: Cimmyt.

He, Z., Xia, X.-C., Chen, X.-M. and Zhuang, Q.-S. (2011) Progress of wheat breeding in China and the future perspective. *Acta Agronomica Sinica*, **37**(2), 202–215 (in Chinese, with English abstract).

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

Huw, J., Lister, D.L., Bower, M.A., Leigh, F.J., Smith, L.M. and Jones, M.K. (2008) Approaches and constraints of using existing landrace and extant plant material to understand agricultural spread in prehistory. *Plant Genet. Resour.* **6**, 98–112.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F. and Orlando, L. (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, **29**, 1682–1684.

Kircher, M., Sawyer, S. and Meyer, M. (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3.

Kistler, L. (2012) Ancient DNA extraction from plants. *Methods Mol. Biol.* **840**, 71–79.

Kistler, L. and Shapiro, B. (2011) Ancient DNA confirms a local origin of domesticated chenopod in eastern North America. *J. Archaeol. Sci.* **38**, 3549–3554.

Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li, S. and Mo, D. (2004) Considering the dating of carbonized wheat at the site of Donghuishan. *Archaeology and Cultural Relics*, **6**, 51–60 (in Chinese).

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, C., Lister, D.L., Li, H., Xu, Y., Cui, Y., Bower, M.A., Jones, M.K. and Zhou, H. (2011) Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. *J. Archaeol. Sci.* **38**, 115–119.

Li, C., Ning, C., Hagelberg, E., Li, H., Zhao, Y., Li, W., Abuduresule, I., Zhu, H. and Zhou, H. (2015) Analysis of ancient human mitochondrial DNA from the Xiaohe cemetery: insights into prehistoric population movements in the Tarim Basin, China. *BMC Genet.* **16**, 78.

Mascher, M., Schuenemann, V.J., Davidovich, U. *et al.* (2016) Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* **48**, 1089–1093.

Miller, C.N., Harper, A.L., Trick, M., Werner, P., Waldron, K. and Bancroft, I. (2016) Elucidation of the genetic basis of variation for stem strength characteristics in bread wheat by Associative Transcriptomics. *BMC Genom.* **17**, 500.

Millward, J.A. (2007) *Eurasian Crossroads: A History of Xinjiang*. New York, NY: Columbia University Press.

Palmer, S.A., Clapham, A.J., Rose, P., Freitas, F.O., Owen, B.D., Beresford-Jones, D., Moore, J.D., Kitchen, J.L. and Allaby, R.G. (2012) Archaeogenomic evidence of punctuated genome evolution in Gossypium. *Mol. Biol. Evol.* **29**, 2031–2038.

Patterson, N., Price, A.L. and Reich, D. (2007) Population structure and eigenanalysis. *PLoS Genet.* **2**, e190.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012) Ancient admixture in human history. *Genetics*, **192**, 1065.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J. and Nieselt, K. (2016) EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60.

Pickrell, J.K. and Pritchard, J.K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967.

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

Ramos-Madrigal, J., Smith, B.D., Moreno-Mayar, J.V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M.T.P. and Wales, N. (2016) Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* **26**, 3195–3201.

Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. and Martin, W. (2002) Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* **3**, 429–441.

Schaefer, N.K., Shapiro, B. and Green, R.E. (2016) Detecting hybridization using ancient DNA. *Mol. Ecol.* **25**, 2398–2412.

Schubert, M., Lindgreen, S. and Orlando, L. (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes*, **9**, 88.

Smith, O., Momber, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V. and Allaby, R.G. (2015) Archaeology. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science*, **347**, 998–1001.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 2725–2729.

Vohr, S.H., Najar, C.F.B.A., Shapiro, B. and Green, R.E. (2015) A method for positive forensic identification of samples from extremely low-coverage sequence data. *BMC Genom.* **16**, 1034.

Wales, N., Madrigal, J.R., Cappellini, E., Baez, A.C., Castruita, J.A.S., Romero-Navarro, J.A., Carøe, C., Ávila-Arcos, M.C., Peñaloza, F. and Moreno-Mayar, J.V. (2016) The limits and potential of paleogenomic techniques for reconstructing grapevine domestication. *J. Archaeol. Sci.* **72**, 57–70.

Wang, J., Luo, M.C., Chen, Z., You, F.M., Wei, Y., Zheng, Y. and Dvorák, J. (2013) Aegilops tauschii single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* **198**, 925–937.

Wang, Y., Yu, H., Tian, C., Sajjad, M., Gao, C., Tong, Y., Wang, X. and Jiao, Y. (2017) Transcriptome association identifies regulators of wheat spike architecture. *Plant Physiol.* **175**, 746–757.

Weiss, C.L., Dannemann, M., Prufer, K. and Burbano, H.A. (2015) Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *eLife*, **4**, e10005.

Wurschum, T., Langer, S.M., Longin, C.F., Korzun, V., Akhunov, E., Ebmeyer, E., Schachschneider, R., Schacht, J., Kazman, E. and Reif, J.C. (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor. Appl. Genet.* **126**, 1477–1486.

Xia, X. and Zhao, Y. (2005) Some latest achievements in research on environment and its evolution in Lop Nur region. *Bull. Nat. Sci. Found. China*, **19**, 196–200.

Xie, M., Yang, Y., Wang, B. and Wang, C. (2013) Interdisciplinary investigation on ancient Ephedra twigs from Gumugou Cemetery (3800b.p.) in Xinjiang region, northwest China. *Microsc. Res. Tech.* **76**, 663–672.

Yang, R., Yang, Y., Li, W., Abuduresule, Y., Hu, X., Wang, C. and Jiang, H. (2014) Investigation of cereal remains at the Xiaohe Cemetery in Xinjiang, China. *J. Archaeol. Sci.* **49**, 42–47.

Zhang, G., Wang, S., Ferguson, D.K., Yang, Y., Liu, X. and Jiang, H. (2017) Ancient plant use and palaeoenvironmental analysis at the Gumugou Cemetery, Xinjiang, China: implication from desiccated plant remains. *Archaeol. Anthrop. Sci.* **9**, 145–152.

Zhao, Z. (2009) Eastward spread of wheat into China – new data and new issues. *Chin. Archaeol.* **9**, 1–9.

Zhou, Y., Chen, Z., Cheng, M. *et al.* (2018) Uncovering the dispersion history, adaptive evolution and selection of wheat in China. *Plant Biotechnol. J.* **16**, 280–291.

Zhuang, Q. (2003) *Chinese Wheat Improvement and Pedigree Analysis.* (Chinese). Beijing, China: China Agriculture Press.

