



# Ancient Mitogenomes Reveal the Origins and Genetic Structure of the Neolithic Shimao Population in Northern China

Jiayang Xue<sup>1,2†</sup>, Wenjun Wang<sup>1,3†</sup>, Jing Shao<sup>4†</sup>, Xiangming Dai<sup>5†</sup>, Zhouyong Sun<sup>4</sup>, Jacob D. Gardner<sup>1</sup>, Liang Chen<sup>6</sup>, Xiaoning Guo<sup>4</sup>, Nan Di<sup>4</sup>, Xuesong Pei<sup>4</sup>, Xiaohong Wu<sup>7</sup>, Ganyu Zhang<sup>1</sup>, Can Cui<sup>1,2</sup>, Peng Cao<sup>1</sup>, Feng Liu<sup>1</sup>, Qingyan Dai<sup>1</sup>, Xiaotian Feng<sup>1</sup>, Ruowei Yang<sup>1</sup>, Wanjing Ping<sup>1,2</sup>, Lizhao Zhang<sup>1</sup>, Nu He<sup>8\*</sup> and Qiaomei Fu<sup>1,2,9\*</sup>

## OPEN ACCESS

### Edited by:

Shaoqing Wen,  
Fudan University, China

### Reviewed by:

Yinqiu Cui,  
Jilin University, China  
Hong-Xiang Zheng,  
Fudan University, China

### \*Correspondence:

Qiaomei Fu  
fuqiaomei@ivpp.ac.cn  
Nu He  
henu@cass.org.cn

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

Received: 31 March 2022

Accepted: 25 April 2022

Published: 27 May 2022

### Citation:

Xue J, Wang W, Shao J, Dai X, Sun Z,  
Gardner JD, Chen L, Guo X, Di N,  
Pei X, Wu X, Zhang G, Cui C, Cao P,  
Liu F, Dai Q, Feng X, Yang R, Ping W,  
Zhang L, He N and Fu Q (2022) Ancient  
Mitogenomes Reveal the Origins and  
Genetic Structure of the Neolithic  
Shimao Population in Northern China.  
*Front. Genet.* 13:909267.  
doi: 10.3389/fgene.2022.909267

<sup>1</sup>Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, China, <sup>2</sup>University of Chinese Academy of Sciences, Beijing, China, <sup>3</sup>Science and Technology Archaeology, National Centre for Archaeology, Beijing, China, <sup>4</sup>Shaanxi Academy of Archaeology, Xi'an, China, <sup>5</sup>Archaeology Institute of National Museum of China, Beijing, China, <sup>6</sup>School of Cultural Heritage, Northwest University, Xi'an, China, <sup>7</sup>School of Archaeology and Museology, Peking University, Beijing, China, <sup>8</sup>Institute of Archaeology, Chinese Academy of Social Sciences, Beijing, China, <sup>9</sup>Shanghai Qi Zhi Institute, Shanghai, China

Shimao City is considered an important political and religious center during the Late Neolithic Longshan period of the Middle Yellow River basin. The genetic history and population dynamics among the Shimao and other ancient populations, especially the Taosi-related populations, remain unknown. Here, we sequenced 172 complete mitochondrial genomes, ranging from the Yangshao to Longshan period, from individuals related to the Shimao culture in northern Shaanxi Province and Taosi culture in southern Shanxi Province, Middle Yellow River basin. Our results show that the populations inhabiting Shimao City had close genetic connections with an earlier population in the Middle Neolithic Yangshao period of northern Shaanxi Province, revealing a mostly local origin for the Shimao Society. In addition, among the populations in other regions of the Yellow River basin, the Shimao-related populations had the closest maternal affinity with the contemporaneous Taosi populations from the Longshan period. The Shimao-related populations also shared more affinity with present-day northern Han populations than with the minorities and southern Han in China. Our study provides a new perspective on the genetic origins and structure of the Shimao people and the population dynamics in the Middle Yellow River basin during the Neolithic period.

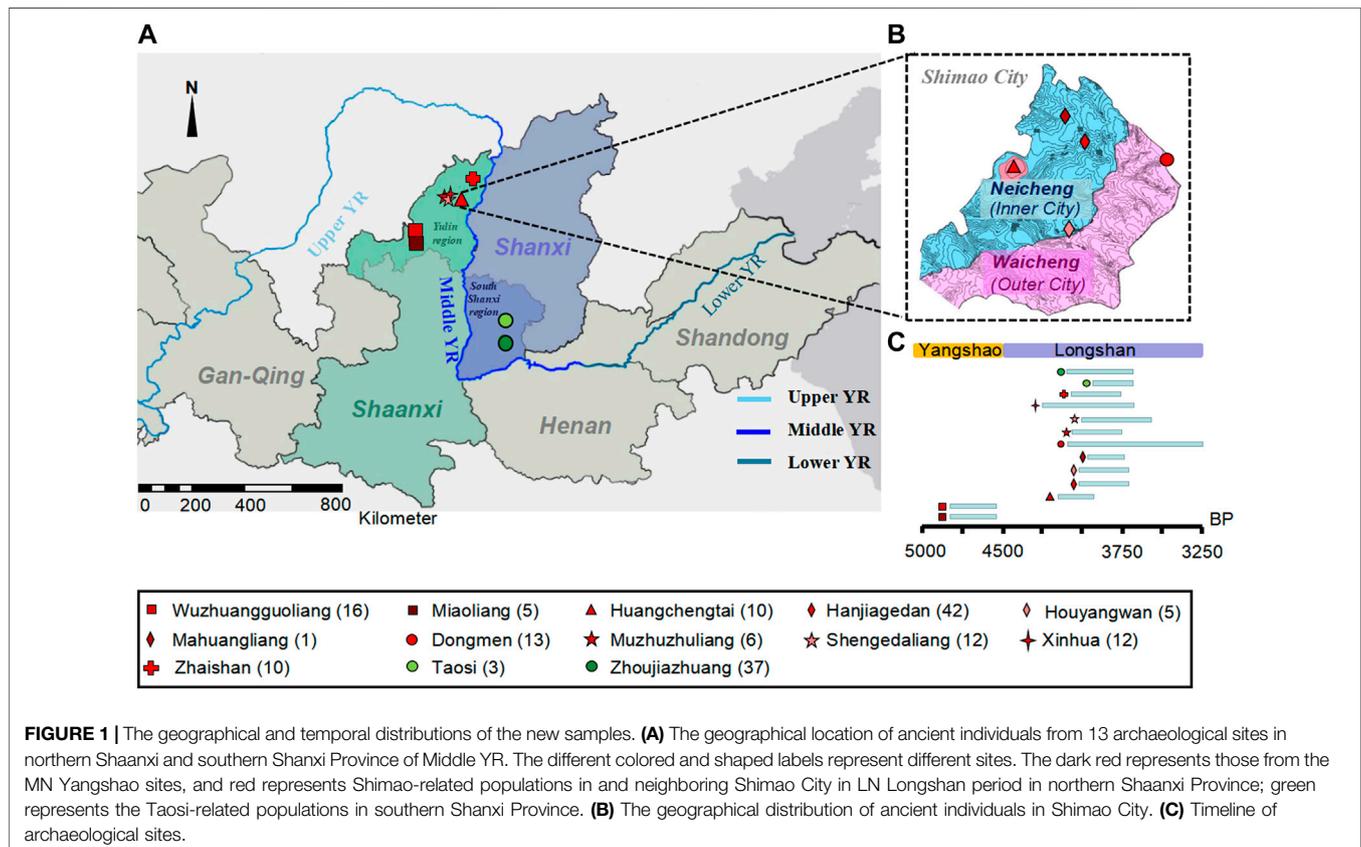
**Keywords:** Shimao, mitochondrial genome, Ancient DNA, Yellow River, Neolithic

**Abbreviations:** BP, before present; CWE, central-west Eurasians; DM, Dongmen; HCT, Huangchengtai; LN, late Neolithic; MN, middle Neolithic; MZZL, Muzhuzhuliang; mtDNA, Mitochondrial DNA; NC, Neicheng; NEAs, Northeastern Asians; PCA, principal component analysis; SEAs, Southeastern Asians; SGDL, Shengedaliang; TS, Taosi; XH, Xinhua; YR, Yellow River; ZS, Zhaishan; ZJZ, Zhoujiazhuang.

## INTRODUCTION

Northern China is a large geographic region that encompasses the Yellow River (YR) basin, in which the residing Neolithic cultures (such as the Yangshao and Longshan cultures) laid an important foundation for the origin of Chinese civilization (Wang, 1989; Dong et al., 2016; Zhou, 2017). The Middle Neolithic (MN) Yangshao period (~7,000–5,000 years before present, BP) was a stage of rapid development and expansion, giving rise to the Majiayao culture (~5,700 BP) in the Upper YR (Yang, 2016), the Daheacun culture in the Middle YR (~5,700 BP) (Xu, 2004), and the Beixin (~5,400 BP) and Dawenkou cultures (~6,000 BP) in the Lower YR (Zhang, 1996; Wang, 2005; Zhang, 2015; Dong et al., 2016; Duan, 2019). This cultural development and expansion coincided with the Holocene Climatic Optimum period in northern China (Hou et al., 2019). In the Late Neolithic (LN) Longshan period (~4,500–3,800 BP), the cultural features in different regions (~4,300 BP in Shaanxi Province; ~4,400 BP in Henan Province; ~4,500 BP in Shanxi Province) of the YR basin varied spatially, increased in social complexity, and formed distinct settlements with different levels of social hierarchy (He, 2004; Chang, 2005; Sun, 2016). The influences of different archaeological cultures on various regions of the YR basin changed dynamically over time, which might have been accompanied by population flow and interaction (Sun et al., 2020a).

The Shimao site (~4,300–3,800 BP), also called “Shimao City”, is considered an important political and religious center during the Middle YR’s Longshan period (~4,500–3,800 BP) (Sun et al., 2013; Rawson, 2017; Sun et al., 2020a). It is currently the largest Neolithic settlement known in China, covering 4 km<sup>2</sup> with a triple structure made of stone-reinforced walls (**Figure 1**), and was selected as one of the world’s top 10 archaeological discoveries in the past decade (Archaeology Institute of America, 2021). The center of Shimao City, Huangchengtai, has many high-grade buildings and relics (Sun et al., 2020a; Sun et al., 2020b). The Neicheng (or “inner city”) surrounds Huangchengtai and consists of multiple grave sites (e.g., Hanjiagedan, Houyangwan, and Mahuangliang). The Dongmen (or “East Gate”) is located along the northeastern wall of Waicheng (or “outer city”) and exhibits complex fortifications (Sun et al., 2020a). According to archaeological records, these distinct sites within Shimao City showed clear differences in social hierarchy and inequality. For example, Hanjiagedan (Sun et al., 2016), Houyangwan (Sun and Shao, 2015), and those closer to Huangchengtai, yielded more high-status graves than Dongmen (Sun and Shao, 2015). Archaeologists named the “Shimao culture” based on artifacts unearthed in Shimao City (Dai, 1998; Sun et al., 2020b). The sites neighboring Shimao City in northern Shaanxi Province, such as the Muzhuzhuliang (Wang et al., 2015), Shengedaliang (Guo et al., 2016), Xinhua (Xing et al., 2002), and Zhaishan sites



**FIGURE 1** | The geographical and temporal distributions of the new samples. **(A)** The geographical location of ancient individuals from 13 archaeological sites in northern Shaanxi and southern Shanxi Province of Middle YR. The different colored and shaped labels represent different sites. The dark red represents those from the MN Yangshao sites, and red represents Shimao-related populations in and neighboring Shimao City in LN Longshan period in northern Shaanxi Province; green represents the Taosi-related populations in southern Shanxi Province. **(B)** The geographical distribution of ancient individuals in Shimao City. **(C)** Timeline of archaeological sites.

(Shao et al., 2021), were all attributed to the Shimao culture. However, the origin of Shimao City was still uncertain. The Shimao culture was considered to have developed from local populations with an influence from surrounding cultures; however, it may have alternatively originated from the migration of populations from the Central Plain or other regions (Dai, 1998; Zhang, 2004; Sun, 2005). In addition, recent studies revealed that the Shimao culture interacted frequently with other regions in the YR basin outside northern Shaanxi Province during the Neolithic period (Guo, 2013; Chen et al., 2016; Chen et al., 2017; Sun et al., 2020b), especially the Taosi culture in southern Shanxi Province of the Middle YR (Yan and He, 2005; Tian and Dai, 2018; Shao, 2020). The links between these two cultures may have been political, economic, cultural, or through shared population connections (Shao, 2020). However, the interactions between the Shimao and Taosi people remain ambiguous from the perspective of archaeology and physical anthropology (Yan and He, 2005; Guo, 2013; Chen et al., 2016; Chen et al., 2017; Tian and Dai, 2018; Sun et al., 2020b; Shao, 2020). Although there were some genomic analyses that included samples from the Shengedaliang (Ning et al., 2020) and Wuzhuangguoliang sites in northern Shaanxi Province (Zhao et al., 2017; Wang et al., 2021), the large-scale genetic affinities among the populations related to the Shimao culture and their predecessors, along with other populations in different regions of the YR basin, is still unclear.

In the current study, we sequenced the complete mitochondrial genomes of 172 samples from various archaeological sites, particularly the individuals related to Shimao and Taosi cultures in the Middle YR. Our research presents large-scale mitogenomic data and new perspectives for exploring the maternal genetic history and dynamics of Shimao-related populations and the populations in the Middle YR basin during the Neolithic period.

## MATERIALS AND METHODS

### Ancient DNA Extraction and Library Preparation

We collected samples from a total of 172 ancient human individuals from 13 sites. We describe their archaeological details in the Supplementary Materials (**Supplementary Table S1**). A total of 172 DNA extractions were obtained from less than 100 mg of bones or dental remains of ancient samples. All ancient DNA work was conducted in the aDNA clean room at the Institute of Vertebrate Paleontology and Paleoanthropology, following strict aDNA standards (Gilbert et al., 2005).

We prepared single-stranded (denoted as “SS”) and double-stranded (denoted as “DS” in **Supplementary Table S1**) libraries and partially treated them with uracil-DNA glycosylase (“UDG”) to remove deaminated cytosine (Kircher et al., 2021; Meyer et al., 2012). Libraries were amplified for 35 cycles using AccuPrime Pfx DNA polymerase under conditions described in a previous study (Dabney and Meyer, 2012). P5 and P7 adapters were added to limit the contamination rate (Kircher et al., 2012). And the

NanoDrop2000 spectrometer was used for monitoring the DNA concentration.

### Ancient DNA Capture and Sequencing

To enrich endogenous ancient DNA from the high levels of background environmental DNA, we used a DNA capture technique (Fu et al., 2013; Fu et al., 2015; Haak et al., 2015). The in-solution capture of the mitochondrial DNA (mtDNA) was accomplished by overlapping probes with DNA fragments and enriching the resulting libraries (Fu et al., 2013). The probes were synthesized based on the human mitochondrial genome.

After enrichment, the Illumina Miseq platform was used to generate  $2 \times 76$  bp paired-end reads. The leeHom software (<https://github.com/greand/leeHom>) was used to trim adapters and merge sequences, with paired-end reads overlapping by at least 11 bp (Gabriel et al., 2014). Sequenced and merged reads with lengths of at least 30 bp were then mapped to the revised Cambridge Reference Sequence version 17 (rCRS, Genbank accession number NC\_012920) for mtDNA, using the same command in the BWA v0.6.1 aligner (arguments used: `-n 0.01` and `-l16500`) (Li and Durbin, 2009; Renaud et al., 2014). We removed duplicate sequences and retained the one with highest mapping quality. After removing sequences with a mapping quality below 30, we constructed the whole mitochondrial sequence (**Supplementary Table S1**).

### Test for Contamination

We evaluated the contamination rate using the ContamMix software and compared the mtDNA fragments with the consensus mitochondrial genome for our newly sampled individuals and 311 present-day world-wide sequences (Fu et al., 2013). We treated the libraries as contaminated if over 4% of the fragments matched with other sequences that are better than the consensus (Fu et al., 2016). For the libraries with substantial contamination (contamination rate > 4%), we excluded them (Briggs et al., 2007; Rohland et al., 2015). Of the 172 new mtDNA samples, 166 of them have lower contamination rates (< 4%, average 0.95%) (**Supplementary Table S1**).

### Kinship Analysis

We treated the mtDNA sequences as kinship-related individuals if they shared identical mitochondrial genomic sequences and were from the same tombs. The Bioedit software (version 7.2.5) was used for testing the kinship. Finally, we found four pairs of sequences with probable matrilineal kinships (Hall, 1999). We excluded the samples in each pair that had relatively lower coverage (**Supplementary Table S1**).

In total, we sequenced the complete mitochondrial genomes of 172 individuals. After removing those with higher contamination rates and kinships, 162 individuals sequenced to between 20.74-fold and 827.53-fold coverage (average 253.95-fold) were used for analysis (**Supplementary Table S1**).

### Haplogroup Analysis

MUSCLE (MUSCLE 3.8.31) and Bioedit software were used to align and edit the complete sequences of mtDNA with rCRS

(Andrews et al., 1999; Edgar, 2004; Weissensteiner et al., 2016). Haplogrep2, built on Phylotree Build 17, was used to call haplogroups for each sample (van Oven, 2015; Weissensteiner et al., 2016) (**Supplementary Table S1**). We grouped all haplogroups absent from the populations in and around Shimao City as “Other”. Since the haplogroups R and N distribute in both East and West Eurasian, we used R<sup>#</sup> and N<sup>#</sup> (such as haplotype R+16189, sub-haplogroups R11 and N9, which were also observed in the Shimao-related populations) to represent the haplogroups carried by East Eurasian in our dataset. The other sub-haplogroups R and N (such as sub-haplogroups R1, R2, N1, N2) found in West Eurasians, were assigned to “Other.”

## Principal Component Analysis and Haplogroup Sharing

In addition, we calculated haplogroup frequencies for every group and conducted a PCA using the built-in function “prcomp” in R (version 4.1.2) software (Venables and Ripley, 2002). We plotted PC1 and PC2 to illustrate the haplogroup differences among populations and explore the maternal genetic relationship among the populations.

We calculated haplogroup sharing as a pairwise matrix of proportions of haplogroups shared among populations. In the matrix, the entries represent the proportions of haplogroups shared in the two populations, which are calculated by summing the shared frequencies of all the same haplogroups between them (Ko et al., 2014; Miao et al., 2021). The proportions of each population were normalized by dividing by the total count and summing to one (Ko et al., 2014).

## Discriminant Analysis of Principal Components

We also used discriminant analysis of principal components (DAPC), which maximizes the inter-variation between groups while minimizing the intra-variation, to show the maternal genetic relationships among ancient populations in the YR basin (Pritchard et al., 2000). We used the DAPC function of the “adegenet” package in R (version 4.1.2) software to conduct a sequence-based DAPC (Jombart et al., 2010).

## Genetic Distance Analysis

The Arlequin software package (version 3.5.2.2) was used to calculate genetic distances ( $F_{ST}$ ) between populations (Excoffier et al., 2007), which were visualized with the package “pheatmap” in R. Generally, a lower  $F_{ST}$  represents a closer maternal genetic relationship between two groups. Heatmaps were also drawn to illustrate the statistical significance of clusters based on  $F_{ST}$ .

## Haplotype Network Construction

To explore the genetic relationship of certain haplotypes in samples, we used DNASP6 (version 6.12.03) and PopArt 1.7 to conduct a median-joining network analysis on all the samples in the same haplogroup (sub-haplogroup or haplotype) dataset (Bandelt et al., 1999; Leigh and Bryant, 2015) and construct the

haplotype network graph. This helped us understand the inflow or diffusion process of a haplotype population.

## RESULTS

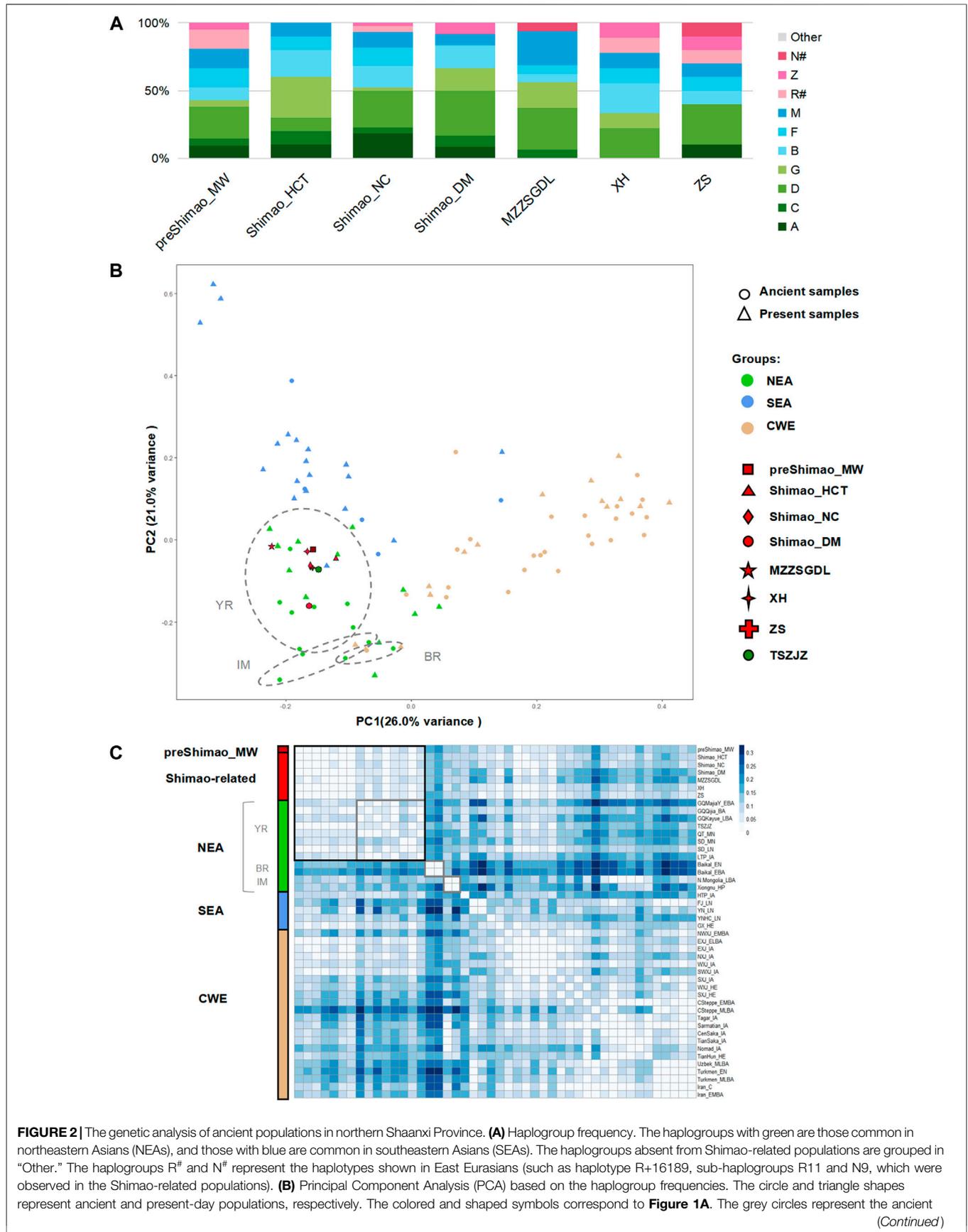
### Sample and Ancient DNA Generation

We captured mitochondrial DNA (mtDNA) from 172 ancient individuals from 13 archaeological sites in the northern Shaanxi and southern Shanxi Provinces of the Middle YR (**Figure 1A**; **Supplementary Table S1**), with dates ranging from 4,836 to 3,253 calibrated BP (cal BP). Removing six individuals with high contamination rates (> 4%) and four individuals owning close relatives (defined as the same mtDNA sequences), resulted in a final dataset of 162 individuals with coverage between 20.74- and 827.53-fold (**Supplementary Table S1**).

Among these new samples, we obtained 21 samples from the Miaoliang and Wuzhuangguoliang sites (referred to as the “preShimao\_MW” group) in northern Shaanxi Province, dating to 4,836–4,530 cal BP in the MN Yangshao period. Additionally, we obtained 91 samples from the LN Longshan period of northern Shaanxi Province, of which 66 were from Shimao City and 35 were from sites neighboring Shimao City (**Supplementary Table S1**). We grouped the individuals in Shimao City based on their archaeological cultures, dates, and geographical locations within Shimao: 10 individuals were excavated from the political and religious center, Huangchengtai site, which we named “Shimao\_HCT” (4,148–3,895 cal BP); 44 individuals were from the Neicheng (or “inner city”), which we grouped as “Shimao\_NC” (3,977–3,699 cal BP) and contained individuals from the Hanjiagedan, Houyangwan, and Mahuangliang sites; and 12 individuals from the Dongmen site in Waicheng (also called the “East Gate in outer city”), which we named “Shimao\_DM” (4,144–3,253 cal BP). The individuals excavated from the Xinhua (XH,  $n = 9$ , 4,231–3,650 cal BP), Muzhuzhuliang (MZZL,  $n = 4$ , 4,082–3,722 cal BP), Shengedaliang (SGDL,  $n = 12$ , 3,969–3,570 cal BP), and Zhaishan (ZS,  $n = 10$ , ~4,050–3,750 BP) sites neighboring Shimao City, we named as the abbreviations of their site names. We integrated MZZL and SGDL into “MZZSGDL” ( $n = 16$ ) for their similar archaeological cultures, locations, and dates, and the small population size ( $n = 4$ ) from the MZZL site.

In addition, we sequenced mtDNA from 40 LN individuals excavated from the “TSZJZ” group (~4,150–3,696 BP) related to Taosi culture (containing Taosi and Zhoujiazhuang sites) in the southern Shanxi Province of the Middle YR.

We also collected 801 previously published mtDNA sequences for the ancient individuals from East and West Eurasia, ranging from the Early Neolithic (EN) to Historic Era (HE). These included populations from Xinjiang (~5,000–500 BP), Gansu and Qinghai Provinces (~5,040–411 BP), Henan Province (~5,500–5,000 BP, Qingtai site), Shandong Province (~9,600–2,000 BP), the Tibetan Plateau (~3,000–100 BP), southern East Asia (~4,600–300 BP), the Baikal River in southern Siberia (~7,123–6,319 BP and ~4,860–3,760 BP), Mongolia (~3,330–2,950 BP and ~2,147–2,007 BP), and the



**FIGURE 2 |** The genetic analysis of ancient populations in northern Shaanxi Province. **(A)** Haplogroup frequency. The haplogroups with green are those common in northeastern Asians (NEAs), and those with blue are common in southeastern Asians (SEAs). The haplogroups absent from Shimao-related populations are grouped in “Other.” The haplogroups R# and N# represent the haplotypes shown in East Eurasians (such as haplotype R+16189, sub-haplogroups R11 and N9, which were observed in the Shimao-related populations). **(B)** Principal Component Analysis (PCA) based on the haplogroup frequencies. The circle and triangle shapes represent ancient and present-day populations, respectively. The colored and shaped symbols correspond to **Figure 1A**. The grey circles represent the ancient (Continued)

**FIGURE 2** | populations from the Yellow River basin (YR), Mongolia and Inner Mongolia (IM), and the Baikal Lake region (BR) in NEA. CWE: Central and Western Eurasian. **(C)** The genetic distance ( $F_{ST}$ ) heatmap of Shimao-related populations and other ancient populations. The different labels and colors correspond to the PCA plot, and different shades of color are used to mark different regional populations. Values with  $F_{ST} = 0.00$  are in white, representing a close genetic relationship. SD\_EN were excluded in the heatmap because of the significantly large genetic distance ( $F_{ST} > 0.10$ ) between them and other populations. The grey squares represent the ancient populations from the Yellow River basin (YR), Mongolia and Inner Mongolia (IM), and the Baikal Lake region (BR) in NEA.

Steppe and West Eurasia (~5,450–1,500 BP). Meanwhile, we also obtained the haplogroup information from individuals in Inner Mongolia (~4,500 BP, Halahaigou site) (**Supplementary Table S2**).

For the present-day populations, we collected 7,641 individuals from northeastern Asians (NEAs, including North Asians and northern East Asians), southeastern Asians (SEAs, Southeast Asians and southern East Asians), and central-west Eurasians (CWEs). Among these populations, 2,102 individuals from China, including 388 Han individuals from northern China and 168 individuals from southern China, which we named “NChina\_Han” and “SChina\_Han”, respectively. We also collected 548 individuals from 16 ethnic minorities, which covers the vast majority of ethnic minorities in China (**Supplementary Table S3**). We served them as different groups following their minorities. The present-day populations also contained the populations in Tibet (“SChina\_Tibet”) and Taiwan (“SChina\_Taiwan”) (Ko et al., 2014; Kang et al., 2016).

## The Mostly Local Genetic Origin of Shimao Populations From Earlier Populations in Northern Shaanxi Province

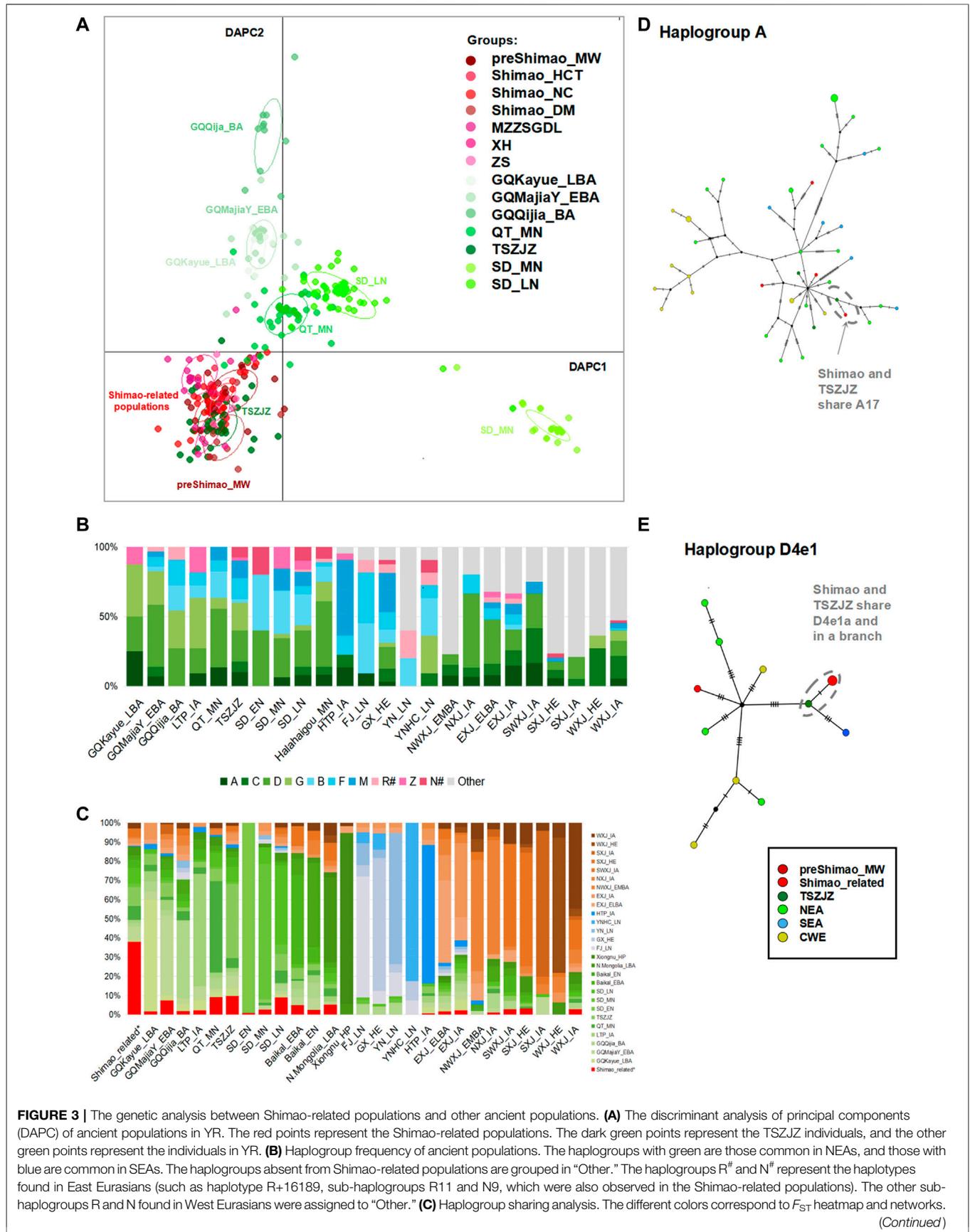
To understand the genetic connection between the LN Shimao populations and the preceding populations in the MN period, we collected 21 individuals from the MN preShimao\_MW sites and 66 individuals from the LN Shimao City in northern Shaanxi Province (**Figure 1**).

The haplogroup analysis found that ancient and present-day NEAs, showed a high proportion of haplogroups A (maximum, 71.43%), C (maximum, 55.00%), D (maximum, 60.00%), and G (maximum, 37.50%) with a north-south declining trend (**Supplementary Figure S1; Supplementary Tables S4, S5**). Haplogroups B (maximum, 36.36%, B4'5), F (maximum, 40.00%), and M (maximum, 83.33%) were common in ancient and present-day SEA and showed a north-south increasing trend (**Supplementary Figure S1; Supplementary Tables S4, S5**). The earlier population in the MN period of northern Shaanxi Province (4,836–4,530 cal BP), preShimao\_MW, carried the haplogroups A (9.52%), C (4.76%), D (23.81%), G (4.76%), B (9.52%, B4'5), F (14.29%), M (14.29%), Z (4.76%), and R<sup>#</sup> (14.29%), and showed a higher proportion of NEA (rather than SEA) haplogroups (**Figure 2A**). Our Principal Component Analysis (PCA) based on haplogroup frequency showed that the PC1 explains population variation from east to west geographically and that PC2 explains the variation from north to south (**Figure 2B**). In general, all the populations are genetically divided into three clusters: NEA, SEA, and CWEs. The preShimao\_MW was distributed among the NEA populations

and clustered with the populations in the YR (**Figure 2B**). In addition, this MN population showed the highest proportion of haplogroup D (23.81%), which was also found in relatively higher proportions in the YR populations (18.18–44.83%) (**Figure 2A, Supplementary Table S4**). The  $F_{ST}$  heatmap based on genetic distance also showed that the preShimao\_MW clustered with the YR populations (**Figure 2C**). Thus, the MN Yangshao populations from northern Shaanxi Province (preShimao\_MW) were more related to the NEA populations in the YR basin than to populations from other regions in East Asia. Although there were no significant genetic affinities between the preShimao\_MW and the EN and MN YR populations ( $F_{ST} > 0.06$ ,  $p < 0.01$  with QT\_MN and SD\_MN; and  $F_{ST} = 0.31$ ,  $p > 0.07$  with SD\_EN), the DAPC shows some overlap between the preShimao\_MW and the QT\_MN from the middle YR (**Figure 3A; Supplementary Table S6**). The same haplotypes G3a2, D5a2a1, and F1a1c were also observed in both preShimao\_MW and QT\_MN, suggesting some connections between them (**Supplementary Figure S2; Supplementary Tables S1, S2**).

For the LN Longshan populations (4,148–3,253 cal BP) in Shimao City, including Shimao\_HCT, Shimao\_NC, and Shimao\_DM, the haplogroup analysis showed that they carried similar haplogroups: A (8.33–18.18%), C (4.55–10.00%), D (10.00–33.33%), G (2.27–30.00%), B (15.91–20.00%, B4'5), and M (8.33–11.36%) (**Supplementary Table S4**). The Shimao populations also displayed higher proportions of NEA dominating haplogroups than SEA dominating haplogroups, and Shimao\_DM (66.67%) showed a higher ratio of NEA dominating haplogroups than Shimao\_HCT (60.00%) and Shimao\_DM (52.27%). Moreover, Shimao\_NC additionally carried haplogroup R<sup>#</sup> (4.55%). Shimao\_NC and Shimao\_DM also had the highest proportions of haplogroup D (27.27–33.33%), similar to preShimao\_MW and most YR populations (**Figure 2A; Supplementary Table S4**). The PCA shows that the three Shimao populations clustered and plot among the NEA populations in the YR region, consistent with the haplogroup analysis results (**Figures 2A,B**). Also, we found that the genetic distances ( $F_{ST}$  values) among the three Shimao populations were all about zero ( $F_{ST} < 0.01$ ,  $p > 0.05$ ), revealing a close genetic affinity among them (**Figure 2C; Supplementary Table S6**). The same haplotypes (B4a4, C4a2, G2a1, and G1c) were found in these three populations, further suggesting a close relationship among them (**Supplementary Figure S2**). These results all suggest that the populations in different regions of the LN Shimao City shared close affinities with each other.

We also explored the genetic connections between the LN Longshan populations (4,148–3,253 cal BP) in Shimao City and the earlier populations (before 4,500 BP) in and outside of



**FIGURE 3 |** The genetic analysis between Shimao-related populations and other ancient populations. **(A)** The discriminant analysis of principal components (DAPC) of ancient populations in YR. The red points represent the Shimao-related populations. The dark green points represent the TSZJZ individuals, and the other green points represent the individuals in YR. **(B)** Haplogroup frequency of ancient populations. The haplogroups with green are those common in NEAs, and those with blue are common in SEAs. The haplogroups absent from Shimao-related populations are grouped in “Other.” The haplogroups R<sup>#</sup> and N<sup>#</sup> represent the haplotypes found in East Eurasians (such as haplotype R+16189, sub-haplogroups R11 and N9, which were also observed in the Shimao-related populations). The other sub-haplogroups R and N found in West Eurasians were assigned to “Other.” **(C)** Haplogroup sharing analysis. The different colors correspond to F<sub>ST</sub> heatmap and networks.

(Continued)

**FIGURE 3** | Median-joining networks of haplotypes A17 (**D**) and D4e1a (**E**) related to ancient northern Chinese populations. The different population groups are shown in different colors that are consistent with those groups in the  $F_{ST}$  heatmap.

northern Shaanxi Province. We found that the MN preShimao\_MW and LN Shimao populations in northern Shaanxi carried similar haplogroups (A, C, D, G, Z, B, F, and M) and that Shimao\_HCT and Shimao\_DM both have the highest proportions of haplogroup D, in common with preShimao\_MW (**Figure 2A**). Some of those haplogroups were absent from the early populations outside Shaanxi Province. For example, the QT\_MN (~5,500–5,000 BP), SD\_EN (~9,600–7,700 BP), and SD\_MN (~5,500–4,600 BP) lacked haplogroup C, and SD\_EN and SD\_MN lacked haplogroup F (**Figure 3B**; **Supplementary Table S4**). The DAPC also indicates that the Shimao populations clustered with the preceding population (preShimao\_MW) in (but not outside) northern Shaanxi Province (**Figure 3A**). Moreover, the three Shimao populations clustered with the preceding MN populations (preShimao\_MW) in the  $F_{ST}$  heatmap and showed the smallest genetic differentiation between them ( $F_{ST} < 0.01$ ,  $p > 0.05$ ), while showing larger  $F_{ST}$  values with the early populations (QT\_MN, SD\_EN, and SD\_MN) outside northern Shaanxi ( $F_{ST} > 0.05$ ) (**Supplementary Table S6**). These results indicate that the LN Longshan populations in Shimao City showed the closest genetic affinity with the earlier MN Yangshao populations (preShimao\_MW, 4,836–4,530 cal BP) in (but not outside) northern Shaanxi Province. This close relationship was also demonstrated by the shared haplotypes between preShimao\_MW and Shimao-related populations, including D4j3 and D4b2b of D4 and haplotypes A+152 + 16362, F1a1c, and R11, and by plotting on the same branches of the median-joining network (**Supplementary Figure S2**). However, we also found some connections between the LN Shimao populations and the earlier QT\_MN from the middle YR. This is supported by the appearance of haplotypes M9a1a1 and M10a1b, which differed in these two populations by only one and four mutations, respectively, in the network analysis (**Supplementary Figure S2**).

Therefore, the populations in different regions of the Longshan period's Shimao City (4,148–3,253 cal BP) shared close affinities with each other and with the preceding MN Yangshao populations (4,836–4,530 cal BP) in (rather than outside) northern Shaanxi Province. The results reveal that the MN Yangshao populations in northern Shaanxi Province were largely not replaced with the foundation of Shimao City, supporting a hypothesis of a mostly local genetic origin for the Shimao people. However, given the shared haplotypes with other YR populations (i.e., the MN Qingtai), we cannot rule out additional genetic contribution from populations outside northern Shaanxi province.

## The Genetic Affinities Among Populations in and Around Shimao City in Northern Shaanxi

The archaeological reports indicated that the LN Longshan sites around Shimao City in northern Shaanxi Province, such as the

MZZSGDL, XH, and ZS, were all related to the Shimao culture. To explore the genetic affinities among the contemporaneous LN populations related to Shimao culture (containing the populations in and around Shimao City), we sequenced 35 new samples from the sites neighboring Shimao City.

These three populations (MZZSGDL, XH, and ZS) around Shimao City primarily carried haplogroups D (22.22–31.25%), B (6.25–22.22%, B4'5), F (6.25–11.11%), and M (10.00–25.00%). Moreover, MZZSGDL also carried haplogroups C (6.25%) and G (18.75%), XH carried haplogroups G (11.11%), Z (11.11%), and R<sup>#</sup> (11.11%), and ZS had haplogroups A (10.00%) and R<sup>#</sup> (10.00%). All three populations displayed the highest proportions of haplogroup D (31.25% in MZZSGDL, 22.22% in XH, and 30.00% in ZS) (**Figure 2A**; **Supplementary Table S4**). In the PCA plot, these populations neighboring Shimao City plot among the NEA populations in YR region and closer to each other (**Figure 2B**). This close relationship was also revealed by the smaller genetic differentiation between them ( $F_{ST} < 0.01$ ) (**Figure 2C**; **Supplementary Table S6**). In the median-joining networks, these populations shared the same branches in haplotype Z3, F2g, and M10a1a1b (**Supplementary Figure S2**). Thus, these three populations neighboring Shimao City were closely related.

In addition, we found that these populations neighboring Shimao City showed the same haplogroups (haplogroups D, B (B4'5), F, and M) as the populations in Shimao City, and some of them share haplotypes (A+152 + 16362, B4a4, D4j3, F2g, and Z3) of the same branch (**Figures 2A,D**; **Supplementary Figure S2**). We also found this close affinity in the DAPC and  $F_{ST}$  heatmap, as well as through the smaller genetic distance between them ( $F_{ST} = 0.00$ ,  $p > 0.05$  in most of them, and  $F_{ST} = 0.04$ ,  $p < 0.05$  between Shimao\_NC and MZZSGDL) (**Figures 2C, 3A**; **Supplementary Table S6**). Among the earlier populations (before 4,500 BP), the individuals neighboring Shimao City showed the closest relationship with the MN Yangshao populations (preShimao\_MW) in (rather than outside) northern Shaanxi Province, similar to the populations in Shimao City, indicated by their shared haplogroups, close distributions in the DAPC, and the smaller  $F_{ST}$  values between them ( $F_{ST} = 0.00$ ,  $p > 0.05$  between preShimao\_MW and XH, ZS; and  $F_{ST} = 0.04$ ,  $p = 0.03$  between preShimao\_MW and MZZSGDL) (**Figures 2A,C, 3A**; **Supplementary Table S6**). Similar with Shimao populations, the populations neighboring Shimao City (such as, MZZSGDL) indicated some slight connections with MN Qingtai supported by the same haplotype M9a1b, which only showed four-mutation differences between them in the network analysis (**Supplementary Figure S2**). Thus, these results show that the populations in and around Shimao City shared close genetic affinities.

Above all, during the LN Longshan period, the populations related to Shimao culture (containing the populations in and around Shimao City, which we call “Shimao-related

populations”) showed close affinities with each other, revealing the extensive connections among the populations not only in but also around Shimao City in northern Shaanxi Province. All Shimao-related populations were shown to have a mostly local genetic origin from the preceding MN Yangshao populations in northern Shaanxi Province.

## The Maternal Affinity Between the Populations Related to Shimao and the Contemporaneous Taosi Culture in the Middle Yellow River

Given the close genetic relationship between the MN Yangshao (4,836–4,530 cal BP) and LN Longshan period (4,231–3,253 cal BP) in northern Shaanxi Province, we then focused on the population interactions between the LN Shimao-related populations and the ancient humans outside northern Shaanxi Province. Previous archaeological studies had shown that stone carvings excavated from Shimao City shared cultural characteristics with the Shang Dynasty (3,500–2,900 BP) in the Central Plain of the Middle YR. Meanwhile, research on pottery from Shimao City found that the Shimao culture was closely associated with the contemporaneous Taosi culture in the southern Shanxi Province of the Middle YR (Shao, 2020; Sun and Shao, 2020). To explore the genetic relationship between Shimao-related populations and the populations in different regions of the YR, we sequenced 40 new individuals from TSZJZ related to the Taosi culture in the southern Shanxi Province of the Middle YR and collected 198 previously published ancient individuals from different regions of the YR.

We found that the Shimao-related populations shared more affinities with those NEA populations in the YR basin. Among those YR populations, the early Bronze Age (EBA) individuals in the Upper YR (GQMajiaY\_EBA) and the LN individuals in the Middle (TSZJZ) and Lower YR (SD\_LN), which date in and after the LN Longshan period (after 4,500 BP), carried higher proportions of haplogroups common in NEAs, such as haplogroups A (6.90–10.00%), C (6.00–7.50%), D (22.50–44.83%), and G (4.00–24.14%). Among these haplogroups, haplogroup D showed the highest proportion in these three YR populations (22.50% in TSZJZ, 26.00% in SD\_LN, and 44.83% in GQMajiaY\_EBA). In addition, they all have the haplogroups B (2.50–22.00%, B4'5), F (6.00–15.00%), and M (3.45–12.50%) (Figure 3B; Supplementary Table S4). These haplogroups were all found in the Shimao-related populations (Figures 2A, 3B). In the DAPC plot, among the Middle YR populations, the Shimao-related populations were closer to the contemporaneous LN Longshan populations (after 4,500 BP) related to the Taosi culture (TSZJZ, 4,150–3,696 cal BP) than to the MN Yangshao populations (~5,500–5,000 BP) from the Qingtai site (QT\_MN, ~5,500–5,000 BP) (Figure 3A). Among the Lower YR populations, the Shimao-related populations were closer to the LN Longshan populations (SD\_LN, after 4,500 BP) than to the EN (SD\_EN, ~9,600–7,700 BP) and MN (SD\_MN, ~5,500–4,600 BP) individuals (Figure 3A). The DAPC suggests that the Shimao-related populations were closer to the contemporaneous LN (after 4,500 BP), rather than the EN and

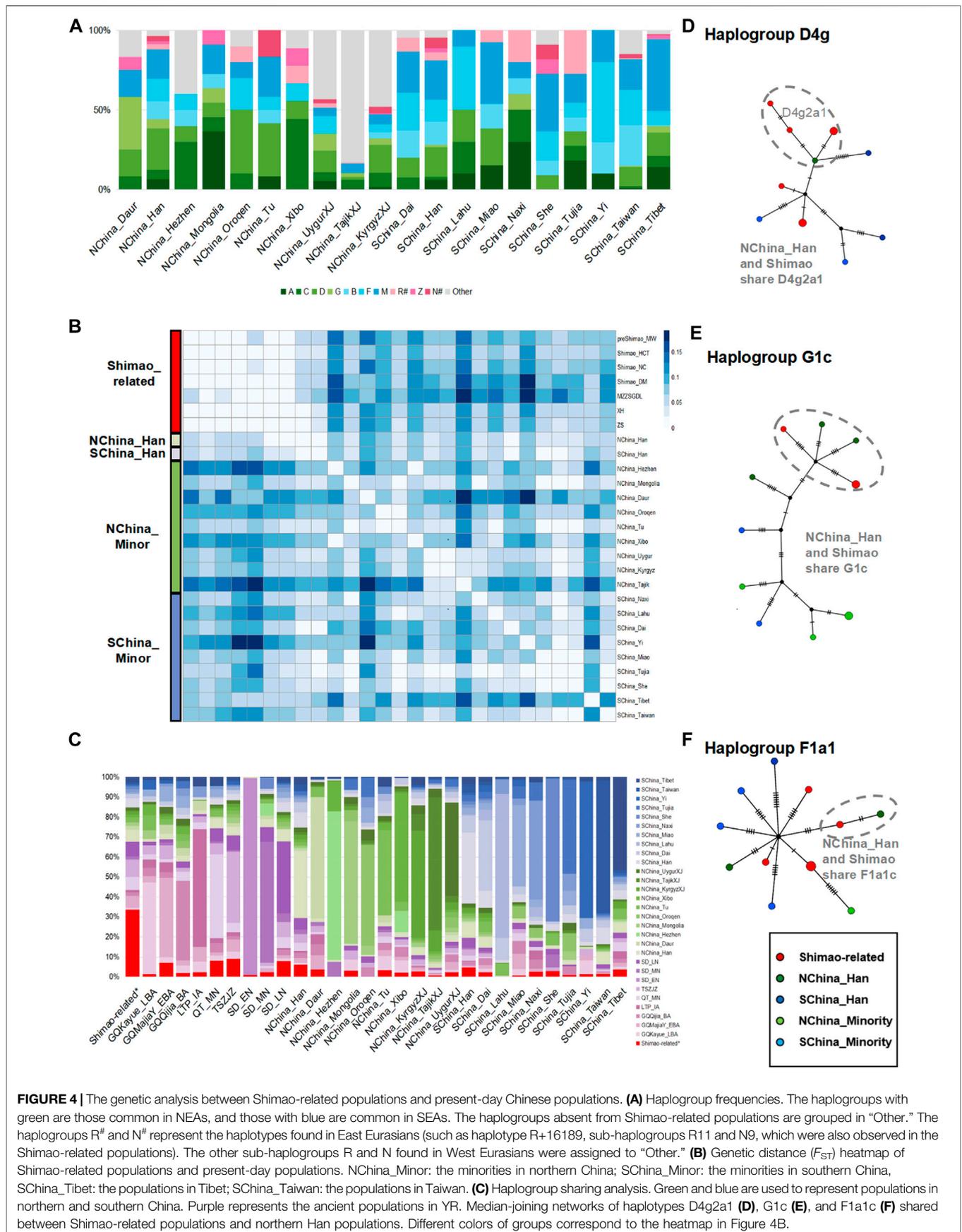
MN (before 4,500 BP) populations outside northern Shaanxi Province in the YR basin. The  $F_{ST}$  results also showed smaller genetic differentiation between Shimao-related individuals and the contemporaneous LN Longshan populations ( $F_{ST} = 0.00$ ,  $p > 0.05$  with TSZJZ;  $F_{ST} < 0.02$ ,  $p > 0.02$  between most Shimao-related populations and SD\_LN;  $F_{ST} = 0.05$ ,  $p = 0.00$  between MZZSGDL and SD\_LN) than with those earlier populations in the YR region ( $F_{ST} = 0.04$ – $0.07$ ,  $p < 0.05$  with QT\_MN;  $F_{ST} = 0.04$ – $0.80$ ,  $p < 0.05$  with SD\_MN; and  $F_{ST} = 0.12$ – $0.49$ ,  $p > 0.05$  with SD\_EN) (Supplementary Table S6). The haplogroup sharing analysis also indicated that the YR populations during the LN shared higher proportions of haplotypes with Shimao-related populations than those in EN and MN (9.80% in TSZJZ, whereas 9.27% in QT\_MN of Middle YR region; 9.05% in SD\_LN, whereas 0.98% in SD\_EN and 2.79% in SD\_MN of Lower YR region) (Figure 3C; Supplementary Table S7). These results also suggest that the Shimao-related populations were closer to LN populations (after 4,500 BP) than the earlier populations (before 4,500 BP) found elsewhere in the YR.

Among the LN populations across the YR, the populations related to the Shimao culture were closest to those related to the Taosi culture ( $F_{ST} = 0.01$ ,  $p > 0.05$  in TSZJZ) (Supplementary Table S6). Similarly, the DAPC results showed Shimao-related populations clearly clustered with TSZJZ to the exclusion of other LN and BA populations (Figure 3A). In addition, the haplogroup sharing analysis also indicated that the Taosi-related individuals shared slightly higher proportions (9.80% in TSZJZ; 9.05% in SD\_LN) of haplotypes with the Shimao-related populations (Figure 3C; Supplementary Table S7). The network results further showed that the haplotypes carried by TSZJZ, such as A17, C4a1a2, C4a2a1, D4b2b, D4e1a, F1a1c, and F2g, were also found in Shimao-related populations and that they shared the same branches (Figures 3D,E; Supplementary Figure S2). These results indicate that, among the populations in other regions of the YR basin, the Shimao-related populations had the closest genetic affinity with the Taosi culture-related populations in southern Shanxi Province.

In all, the ancient individuals related to the Shimao culture in the LN Longshan period from northern Shaanxi Province shared more maternal relationships with the contemporaneous (but not earlier) populations in the YR region outside northern Shaanxi Province. Among these LN Longshan populations, those related to Shimao culture shared the closest relationship with those related to Taosi culture in the Middle YR. These results demonstrate the strong and extensive population interactions during the LN Longshan period, not only within the northern Shaanxi Province but also between northern Shaanxi and southern Shanxi Provinces.

## The Genetic Relationship Between Shimao-Related Populations and Present-Day Humans

To explore the genetic relationships between Shimao-related populations and present-day humans, we compared their genetic affinities including the ethnic minorities (such as Daur, Mongolia, Dai, Miao, etc.), Han populations (“NChina\_Han” and



**FIGURE 4 |** The genetic analysis between Shimao-related populations and present-day Chinese populations. **(A)** Haplogroup frequencies. The haplogroups with green are those common in NEAs, and those with blue are common in SEAs. The haplogroups absent from Shimao-related populations are grouped in “Other.” The haplogroups R<sup>#</sup> and N<sup>#</sup> represent the haplotypes found in East Eurasians (such as haplotype R+16189, sub-haplogroups R11 and N9, which were also observed in the Shimao-related populations). The other sub-haplogroups R and N found in West Eurasians were assigned to “Other.” **(B)** Genetic distance ( $F_{ST}$ ) heatmap of Shimao-related populations and present-day populations. NChina\_Minor: the minorities in northern China; SChina\_Minor: the minorities in southern China, SChina\_Tibet: the populations in Tibet; SChina\_Taiwan: the populations in Taiwan. **(C)** Haplogroup sharing analysis. Green and blue are used to represent populations in northern and southern China. Purple represents the ancient populations in YR. Median-joining networks of haplotypes D4g2a1 **(D)**, G1c **(E)**, and F1a1c **(F)** shared between Shimao-related populations and northern Han populations. Different colors of groups correspond to the heatmap in Figure 4B.

“SChina\_Han”), and the populations in Tibet and Taiwan of China.

Among these present-day populations in China, the Han populations carried both the NEA dominating haplogroups A (5.95–6.46%), C (1.79–5.94%), D (19.05–25.84%), and G (1.19–5.94%) and SEA dominating haplogroups B (11.37–14.88%, B4'5), F (13.69–13.95%), and M (18.60–24.40%), and showed the highest proportion of haplogroup D (19.05%–25.84%), consistent with the Shimao-related populations (**Figure 4A; Supplementary Table S5**). The genetic distance analysis also showed that the Shimao-related populations (such as, Shimao\_HCT, XH, ZS) were closer to Han populations ( $F_{ST} < 0.03$ ,  $p > 0.06$  in NChina\_Han;  $F_{ST} < 0.04$ ,  $p > 0.06$  in SChina\_Han) than other present-day minority populations, including those in Tibet and Taiwan (**Figure 4B; Supplementary Table S6**). The haplogroup sharing analysis showed that the Shimao-related populations shared higher proportions of haplotypes with the Han (NChina\_Han, 6.04%; SChina\_Han, 4.70%) than with the other present-day populations (0.00–3.65%) (**Figure 4C; Supplementary Table S8**). Additionally, the Shimao-related populations (such as, Shimao\_HCT, XH, ZS) were genetically closer to northern Han populations (NChina\_Han,  $F_{ST} < 0.03$ ,  $p > 0.06$ ) than the southern Han populations (SChina\_Han,  $F_{ST} > 0.03$ ,  $p > 0.03$ ) (**Figure 4B; Supplementary Table S6**). Moreover, the haplogroup sharing analysis also demonstrated that the northern Han population shared higher proportions (NChina\_Han, 6.04%) of haplotypes with Shimao-related populations than southern Han populations (SChina\_Han, 4.70%) (**Figure 4C; Supplementary Table S8**). The network analysis also indicated that the Shimao-related populations and northern Han populations shared multiple haplotypes (D4g2a1, G1c, and F1a1) between them (**Figures 4D–F**). Thus, we conclude that the Shimao-related populations were closer to Han populations in northern China than to the minorities and Southern Han populations in China.

To explore which ancient populations had the closest genetic to the northern Han in China, we compared the affinities of our new Shimao-related populations and the ancient individuals in other regions of China to the northern Han. These include the EBA (GQMajiaY\_EBA), BA (GQQijia\_BA), late BA (GQKayue\_LBA), and IA (LTP\_IA) individuals from Gansu-Qinghai Province; the MN individuals (QT\_MN) from Henan Province; the LN (TSZJZ) individuals from southern Shanxi Province; and the EN, MN, and LN individuals in Shandong Province (SD\_EN, SD\_MN, SD\_LN). We found that the haplogroups A, C, D, G, Z, B (B4'5), F, M, and R<sup>#</sup> were observed in the Shimao-related populations, LN Shandong individuals, and NChina\_Han, whereas some of those haplogroups were absent from the other populations (**Figures 2A, 3B, 4A**). For example, TSZJZ lacked haplogroup R<sup>#</sup>, SD\_MN lacked haplogroups C, F, and R<sup>#</sup>, and QT\_MN lacked haplogroups C, R<sup>#</sup>, and Z (**Figure 3B; Supplementary Table S4**). Moreover, most of the Shimao-related populations (22.20–33.30%), SD\_LN (26.00%), and NChina\_Han (25.84%) carried the highest proportions of haplogroup D (**Supplementary Tables S4, S5**). The haplogroup sharing analysis showed that

NChina\_Han shared the highest proportions of haplotypes with Shimao-related populations (6.04%) compared to QT\_MN (4.09%), TSZJZ (4.18%), and SD\_LN (5.75%) (**Figure 4C; Supplementary Table S8**). The genetic distance analysis also showed that NChina\_Han shared the closest genetic affinities with the Shimao-related populations ( $F_{ST} = 0.02$ ,  $p = 0.10$  in XH and ZS;  $F_{ST} = 0.03$ ,  $p = 0.06$  in Shimao\_HCT) compared to QT\_MN ( $F_{ST} = 0.03$ ,  $p = 0.00$ ); TSZJZ ( $F_{ST} = 0.05$ ,  $p = 0.00$ ); SD\_LN ( $F_{ST} = 0.06$ ,  $p = 0.00$ ) (**Supplementary Table S6**). Thus, the Han populations in northern China shared more affinity with the Shimao-related populations than with the other published ancient individuals in China.

We found that the populations in and around Shimao City were closer to the northern Han Chinese populations than to the southern Han Chinese and minority populations. Compared to the other ancient individuals in China, we also found that these Han populations in northern China were closer to the Shimao-related populations.

## DISCUSSION

Archeological research on Shimao City and the corresponding cultures reveal its importance, particularly in the Longshan period (~4,500–3,800 BP), as a crucially political and religious center in northern Shaanxi Province of the Middle YR (Sun et al., 2020a). However, the genetic origins of the Shimao people rested on the relationships between the populations related to the Shimao culture and the local populations in the preceding MN Yangshao period, along with the other ancient populations (especially those contemporaneous) in the YR basin outside northern Shaanxi Province. In the current study, we presented a large-scale dataset of ancient mitochondrial genomes from the ancient populations in northern Shaanxi and southern Shanxi Province of the Middle YR, especially those related to the Shimao and contemporaneous Taosi cultures. Through our new study, we have characterized the genetic structure and population dynamics of the Shimao-related populations and how populations changed from the MN Yangshao to the LN Longshan period through the present day.

First, previous genomic research on the populations in the Henan Province of the Middle YR found distinct genetic composition changes from the MN Yangshao period (~5,550–5,050 BP) to the LN Longshan period (~4,275–3,844 BP), with the latter having more ancestry from the south (Ning et al., 2020). Similarly, the ancient populations in the Shandong region of the Lower YR also exhibited changes in their mitochondrial genomes from the Yangshao (~5,500–4,600 BP) to Longshan (after 4,500 BP) period (Liu et al., 2021). In contrast, our results indicate that the populations in the LN Longshan period related to the Shimao culture (4,231–3,253 cal BP) in northern Shaanxi Province of Middle YR showed the closest genetic connections with their earlier populations in the MN Yangshao period (Wuzhuangguoliang and Miaoliang sites, 4,836–4,530 cal BP). This is consistent with archaeological studies showing that the relics excavated in the Longshan period maintained the features

of those in the Yangshao culture in this region, differentiating them from the Central Plain (Xing et al., 2002).

The origin of Shimao City was also uncertain (Sun et al., 2020a). Some researchers believed it was developed by the local populations and influenced by surrounding archaeological cultures (Gong, 2018), while some hypothesized that populations migrated from the southern region of the Middle YR and built Shimao City (Zhang, 2004; Duan and Dong, 2018). Interestingly, the close genetic connections between Shimao populations in the LN Longshan period and the preceding populations in the MN Yangshao period (Wuzhuangguoliang and Miaoliang sites, 4,836–4,530 cal BP) in our study support a mostly local origin for Shimao and being primarily developed by the Yangshao populations before 4,500 BP in northern Shaanxi Province. However, we also find evidence for some connections between the populations in (MN preShimao\_MW and LN Shimao) and outside (Qingtai in Central Plain) of northern Shaanxi Province. Thus, despite the close connection between Shimao and its MN predecessors in northern Shaanxi Province, we cannot exclude the possibility of additional genetic contribution from the middle YR region. Future studies will clarify these relationships.

Second, the archaeological studies on potteries, tiles, and other relics unearthed from Shimao City revealed the populations in Huangchengtai was the core palaces with high hierarchy, while the populations in the Dongmen of Waicheng were mainly from sacrificial pits, which were the sacrificed victims with low social status (Sun and Shao, 2020). However, our findings revealed that these populations in different regions of Shimao City showed close genetic affinities, though they had different levels of inferred social status. Those populations in Xinhua, Muzhuzhuliang, Shengedaliang, and Zhaishan sites related to Shimao culture had varying levels of social hierarchy and also showed a close relationship with the ancestors of the Shimao people, indicating the extensive connections among the Shimao-related populations. Thus, although the LN Longshan populations in and around Shimao showed varying levels of social hierarchy, they also shared close genetic connections with each other, consistent with their proposed cultural connections (Sun et al., 2020a).

Moreover, the physical anthropology of the individuals in Dongmen suggested that their skulls were similar to those related to the Lower Xiajiadian culture (~4,150–3,590 BP) in Inner Mongolia (Chen et al., 2016). Published mitochondrial haplogroups from individuals at the Erdaojingzi site (WLR\_LN in Ning et al., 2020), related to the Lower Xiajiadian culture, included B5b1a, A22, N9a1, which were all absent from the individuals at the Dongmen of Shimao City in our study. Stone carvings found in Shimao City (such as, in the center palace of Huangchengtai and Dongmen) were also observed in the Xinglongwa (~8,200–7,400 BP) and Zhaobaogou (~7,350–6,420 BP) culture related sites in Inner Mongolia (Lv, 1960; Wang, 1993; Sun and Shao, 2020). We didn't find a closely maternal affinity between the populations in Dongmen and Inner Mongolia; however, more ancient genomes from Inner Mongolia are needed to confirm the genetic affinity between them.

We then focused on the genetic affinity between Shimao-related populations and the ancient populations in other regions of the YR. Our results found that the Shimao-related populations in the LN Longshan period shared more genetic relationships with the contemporaneous LN populations in the YR region, rather than those from earlier periods. Among these YR populations in the LN Longshan period, the populations related to Shimao culture showed the closest affinity with the populations related to the Taosi culture in the Middle YR's southern Shanxi Province. Similarly, Xu (2014) conducted comparative studies of jade burial, painting and violence excavated from Shimao and Taosi cities, and showed that there were similarities in culture, and interaction in the economy, culture, and populations in these two vital cities (Xu, 2014), which were in accordance with our new findings.

Finally, we found that the LN Longshan Shimao-related populations in northern Shaanxi were closer to the present-day Han Chinese (especially the northern Han Chinese population) than to the minorities in China. A previous genetic study found that the MN Qingtai population also contributed to the Han in northern China. We found that northern Han Chinese populations shared more affinity with our Shimao-related populations than with the ancient Qingtai individuals. According to the recent census reports of the Chinese population and related historical studies, most of the humans in present-day northern Shaanxi were Han Chinese (Zhou, 2015; National Bureau of Statistics, 2021). Meanwhile, the genetic research on Han Chinese populations showed that the Han in northern China were identical and that the north-south genetic divergence was the major difference among Han populations (Li et al., 2019; He et al., 2022). Thus, our results revealed the close genetic affinities between the MN Yangshao (4,830–4,530 cal BP) and LN Longshan (4,231–3,253 cal BP) populations in northern Shaanxi Province and that they contributed to the present-day northern Han populations to a certain extent.

In summary, we conclude that the Shimao-related populations in the LN Longshan period are closely related to their predecessors in the MN Yangshao period in northern Shaanxi Province, revealing a mostly local origin for Shimao City. In addition, compared to other LN populations, the Shimao-related populations were genetically closer to the contemporaneous Taosi population in southern Shanxi Province, reflecting strong interactions in both regions of the Middle YR in the LN. As for their relationship with present-day people, the Shimao-related populations were genetically closer to northern Han population. Our study might provide a perspective for understanding the genetic affinities and population dynamics of Shimao-related populations in the Middle YR basin during the Neolithic period. Further studies with ancient genomic data will aim to test for more complex patterns of admixture and social organization in this region.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Genome Warehouse in

National Genomics Data Center (National Genomics Data Center Members and Partners (CNCB-NGDC), 2020; Chen et al., 2021), Beijing Institute of Genomics (China National Center for Bioinformatics), Chinese Academy of Sciences, accession number PRJCA009290 (<https://bigd.big.ac.cn/gwh>). Other datasets for this study can be found in the **Supplementary Material**.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

QF designed and supervised the research project. QF, WW managed the project. ZS, LC, XG, ND, XP, XW, GZ, CC, LZ, and NH assembled archaeological materials and dating. QF, PC, RY, FL, XF, QD, and WP, performed or supervised wet laboratory work. QF, XF did the data processing and quality control. JX, WW and GZ, did the primary data analysis for this manuscript. JX, WW, JS, XD, JG, and QF, wrote the manuscript. All authors discussed, critically revised, and approved the final version of the manuscript.

## REFERENCES

- Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowers, R. N., Turnbull, D. M., and Howell, N. (1999). Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA. *Nat. Genet.* 23 (2), 147. doi:10.1038/13779
- Archeology Institute of America (2021). Top 10 Discoveries of the Decade. Available at: <https://www.archaeology.org/issues/406-2101/features/9325-china-shimao-neolithic-city>. (Accessed March 22, 2022).
- Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining Networks for Inferring Intraspecific Phylogenies. *Mol. Biol. Evol.* 16 (1), 37–48. doi:10.1093/oxfordjournals.molbev.a026036
- Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., et al. (2007). Patterns of Damage in Genomic DNA Sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* 104 (37), 14616–14621. doi:10.1073/pnas.0704665104
- Chang, H. Y. (2005). Preliminary Study on the Process of Social Complication from Longshan Period to Early Erlitou Period. master's thesis. Chengdu (Sichuan): Sichuan University.
- Chen, L., Sun, Z. Y., and Shao, J. (2017). Research on Neolithic Skeletons from Houyangwan Site of Shimao City in Shenmu County, Shaanxi Province. *Xi Bu Kao Gu.* 12 (03), 263–273.
- Chen, L., Xiong, J. X., Shao, J., and Sun, Z. Y. (2016). The Research of the Forensic Analysis of Skulls from the Sacrificial Pits Located in Shimao Site, Shaanxi Province. *Archaeol. Cult. Relics.* 37 (04), 134–142.
- Chen, M. L., Ma, Y. K., Wu, S., Zheng, X. C., Kang, H. G., Sang, J., et al. (2021). Genome Warehouse: A Public Repository Housing Genome-Scale Data. *Genom. Proteom. Bioinform.* 19 (4), 584–589. doi:10.1016/j.gpb.2021.04.001
- Dabney, J., and Meyer, M. (2012). Length and Gc-Biases during Sequencing Library Amplification: a Comparison of Various Polymerase-Buffer Systems with Ancient and Modern Dna Sequencing Libraries. *BioTechniques* 52 (2), 87–94. doi:10.2144/000113809

## FUNDING

This work was supported by the Chinese Academy of Sciences (CAS) and the Ministry of Finance of the People Republic of China (YSBR-019 and XDB26000000), National Natural Science Foundation of China (41925009, T2192950), National Social Science Fund of China (19ZDA232), the “Research on the Roots of Chinese Civilization” program of Zhengzhou University (XKZDJC202006), the Howard Hughes Medical Institute (grant no. 55008731), the Tencent Foundation (through the XPLOER PRIZE). XW was supported by the Key National Social Science Foundation of China (no. 16ZDA144).

## ACKNOWLEDGMENTS

We would like to thank the archaeological teams from Shaanxi, Archaeology Institute of National Museum of China, and Archaeology Institute of Chinese Academy of Social Sciences.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.909267/full#supplementary-material>

- Dai, X. M. (1998). Evolution of Neolithic Cultural Pattern in the Yellow River Basin. *Acta Archaeologica Sinica.* 63 (04), 389–418.
- Dong, G. H., Liu, F. W., Yang, Y. S., Wang, L., and Chen, F. H. (2016). Cultural Expansion and its Influencing Factors during Neolithic Period in the Yellow River Valley, Northern China. *Chin. J. Nat.* 38 (04), 248–252. doi:10.3969/j.issn.0253-9608.2016.04.003
- Duan, T. J., and Dong, X. L. (2018). On the Nature of Remains in Shimao Site in Northern Shaanxi and Corresponding in the Perspective of *Li* Spectrum. *Res. China's Front. Archaeol.* 17 (02), 218–236.
- Duan, X. Q. (2019). Study on Painted Pottery of Gansu and the Development of Prehistoric Painted Potter. *J. Northwest Univ. Natl.* 41 (06), 6–13. doi:10.14084/j.cnki.cn62-1185/c.2019.06.002
- Edgar, R. C. (2004). MUSCLE: a Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinforma.* 5, 113. doi:10.1186/1471-2105-5-113
- Excoffier, L., Laval, G., and Schneider, S. (2007). Arlequin (Version 3.0): an Integrated Software Package for Population Genetics Data Analysis. *Evol. Bioinform Online* 1, 47–50. doi:10.1143/JJAP.34.L418
- Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., et al. (2016). The Genetic History of Ice Age Europe. *Nature* 534 (7606), 200–205. doi:10.1038/nature17993
- Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., et al. (2015). An Early Modern Human from Romania with a Recent Neanderthal Ancestor. *Nature* 524, 216–219. doi:10.1038/nature14558
- Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., et al. (2013). DNA Analysis of an Early Modern Human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* 110 (6), 2223–2227. doi:10.1073/pnas.1221359110
- Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing Ancient DNA Studies. *Trends Ecol. Evol.* 20 (10), 541–544. doi:10.1016/j.tree.2005.07.005
- Gong, Q. M. (2018). The Important Harvest of Shaanxi Prehistoric Archaeology in the New Century. *Wenbo.* 35 (05), 31–50.
- Guo, W. (2013). “The Communication between Northern China and Eurasian Steppe During the Longshan Period From the Stone Figures at Shimao Site.

- Available at: <https://kaogu.cn/html/cn/xueshuyanjiiu/yanjiuxinlun/julouyuchengshikaog/2013/1025/33681.html> (Accessed March 22, 2022).
- Guo, X. N., Wang, W. L., Kang, N. W., Qu, F. M., and Chen, L. (2016). Archaeology Survey of Excavation at the Shengedaliang Site, Shenmu County. *Archaeol. Cult. Relics* 37 (04), 34. doi:10.3969/j.issn.1000-7830.2016.04.004
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., et al. (2015). Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe. *Nature* 522, 207–211. doi:10.1038/nature14317
- Hall, T. A. (1999). Bioedit: a User-Friendly Biological Sequence Alignment Editor and Analysis. *Nucl. Acids Symp.* 41 (41), 95–98.
- He, G. L., Wang, M. G., Li, Y. X., Zou, X., Yeh, H. Y., Tang, R. K., et al. (2022). Fine-scale North-to-south Genetic Admixture Profile in Shaanxi Han Chinese Revealed by Genome-wide Demographic History Reconstruction. *J. Syst. Evol.* doi:10.1111/jse.12715
- He, N. (2004). Review on the Research History of Taosi Culture. *Anc. Civiliz.* 3 (00), 54–86.
- Hou, G. L., Xu, C. J., Lv, C. Q., Chen, Q., and Lan, C. Z. (2019). Environmental Background of Yangshao Culture Expansion in Holocene. *Geogr. Res.* 38 (02), 437–444. doi:10.11821/dlyj020171214
- Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant Analysis of Principal Components: a New Method for the Analysis of Genetically Structured Populations. *BMC Genet.* 11, 94. doi:10.1186/1471-2156-11-94
- Kang, L., Zheng, H.-X., Zhang, M., Yan, S., Li, L., Liu, L., et al. (2016). MtDNA Analysis Reveals Enriched Pathogenic Mutations in Tibetan Highlanders. *Sci. Rep.* 6, 31083. doi:10.1038/srep31083
- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double Indexing Overcomes Inaccuracies in Multiplex Sequencing on the Illumina Platform. *Nucleic Acids Res.* 40 (1), e3. doi:10.1093/nar/gkr771
- Ko, A. M.-S., Chen, C.-Y., Fu, Q., Delfin, F., Li, M., Chiu, H.-L., et al. (2014). Early Austronesians: into and Out of Taiwan. *Am. J. Hum. Genet.* 94 (3), 426–436. doi:10.1016/j.ajhg.2014.02.003
- Leigh, J. W., and Bryant, D. (2015). Popart: Full-feature Software for Haplotype Network Construction. *Methods Ecol. Evol.* 6, 1110–1116. doi:10.1111/2041-210X.12410
- Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, Y.-C., Ye, W.-J., Jiang, C.-G., Zeng, Z., Tian, J.-Y., Yang, L.-Q., et al. (2019). River Valleys Shaped the Maternal Genetic Landscape of Han Chinese. *Mol. Biol. Evol.* 36 (8), 1643–1652. doi:10.1093/molbev/msz072
- Liu, J., Zeng, W., Sun, B., Mao, X., Zhao, Y., Wang, F., et al. (2021). Maternal Genetic Structure in Ancient Shandong between 9500 and 1800 Years Ago. *Sci. Bull.* 66 (11), 1129–1135. doi:10.1016/j.scib.2021.01.029
- Lv, Z. E. (1960). Archaeological Survey of Linxi Site, Inner Mongolia. *Acta Archaeol. Sin.* 25 (1), 9–23.
- Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338 (6104), 222–226. doi:10.1126/science.1224344
- Miao, B., Liu, Y., Gu, W., Wei, Q., Wu, Q., Wang, W., et al. (2021). Maternal Genetic Structure of a Neolithic Population of the Yangshao Culture. *J. Genet. Genomics* 48 (08), 746–750. doi:10.1016/j.jgg.2021.04.005
- National Bureau of Statistics (2021). *Major Figures on 2020 Population Census of China*. Beijing: China Statistical Press.
- National Genomics Data Center Members and Partners (CNGB-NGDC) (2020). Database Resources of the National Genomics Data Center in 2020. *Nucleic Acids Res.* 48 (D1), D24–D33. doi:10.1093/nar/gkz913
- Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., et al. (2020). Ancient Genomes from Northern China Suggest Links between Subsistence Changes and Human Migration. *Nat. Commun.* 11 (1), 2700. doi:10.1038/s41467-020-16557-2
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155 (2), 945–959. doi:10.1093/genetics/155.2.945
- Rawson, J. (2017). Shimao and Erlitou: New Perspectives on the Origins of the Bronze Industry in Central China. *Antiquity* 91 (355), E5. doi:10.15184/aqy.2016.234
- Renaud, G., Stenzel, U., and Kelso, J. (2014). leeHom: Adaptor Trimming and Merging for Illumina Sequencing Reads. *Nucleic Acids Res.* 42 (18), e141. doi:10.1093/nar/gku699
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial Uracil-DNA-Glycosylase Treatment for Screening of Ancient DNA. *Phil. Trans. R. Soc. B* 370 (1660), 20130624. doi:10.1098/rstb.2013.0624
- Shao, J. (2020). A Comparative Study of Shimao Site and Taosi Site. *Archaeology* 66 (05), 65.
- Shao, J., Pei, X. S., Di, N., Yuan, Y., Zhao, K., and He, C. L. (2021). Preliminary Report on the Excavation of the Residential Area at the Miaoyan Locality of the Zhaishan Site in Fugu County, Shaanxi Province. *Wenbo* 38 (5), 15–28+2+113. doi:10.3969/j.issn.1000-7954.2021.05.003
- Sun, Z. Y. (2005). Review of Xinhua Culture. *Archaeol. Cult. Relics* 26 (03), 40. doi:10.3969/j.issn.1000-7830.2005.03.006
- Sun, Z. Y., Shao, J., Shao, A. D., Zhao, X. H., Yang, G. Q., Kang, N. W., et al. (2016). Archaeology Survey of Excavations to Hanjiagedan Locus of the Shimao Site in Shenmu County, Shaanxi Province. *Archaeol. Cult. Relics* 37 (04), 14. doi:10.3969/j.issn.1000-7830.2016.04.002
- Sun, Z. Y., and Shao, J. (2015). Archaeology Survey of Trial Excavations to Houyangwan and Hujiawa Localities of the Shimao Site in Shenmu County, Shaanxi Province. *Archeology* 61 (05), 60.
- Sun, Z. Y., Shao, J., and Di, N. (2020a). Archaeological Discovery and Research Synthesis of Shimao Site. *Cult. Relics Central China* 44 (01), 39–62.
- Sun, Z. Y., Shao, J., and Di, N. (2020b). Nomenclature, Range and Age of Shimao Culture. *Archaeology* 66 (8), 101–108.
- Sun, Z. Y., and Shao, J. (2020). Research on the Unearthed Stone Sculptures from the Large Foundation of Huangchengtai at Shimao Site. *Archaeol. Cult. Relics* 41 (04), 40. doi:10.3969/j.issn.1000-7830.2020.04.005
- Sun, Z. Y., Shao, J., Shao, A. D., Kang, N. W., Qu, F. M., and Liu, X. M. (2013). Shimao Site in Shenmu County, Shaanxi. *Archaeology* 59 (07), 15.
- Sun, Z. Y. (2016). Study on the Formation Process of Social Complexity in Northern China in the Third Millennium BC. *Archaeol. Cult. Relics* 37 (04), 70–79. doi:10.3969/j.issn.1000-7830.2016.04.008
- Tian, W., and Dai, X. M. (2018). Archaeology Survey on the Excavation of Zhoujiazhuang Site, Jiangxian County, Shanxi Province in 2013. *Archaeology* 64 (01), 28.
- van Oven, M. (2015). PhyloTree Build 17: Growing the Human Mitochondrial DNA Tree. *Forensic Sci. Int. Genet. Suppl. Ser.* 5, e392–e394. doi:10.1016/j.fsigs.2015.09.155
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer.
- Wang, C. C., Yeh, H. Y., Popov, A. N., Zhang, H. Q., Matsumura, H., Sirak, K., et al. (2021). Genomic Insights into the Formation of Human Populations in East Asia. *Nature* 591, 413–419. doi:10.1038/s41586-021-03336-2
- Wang, F. (1989). On the Position of the Yellow River Basin in the Origin of Chinese Civilization - Also on the Cause of the Establishment of Xia Dynasty. *J. Renmin Univ. China* 3 (03), 95–102.
- Wang, G. (1993). Stone Carvings with Portrait in Xinglongwa Culture. *China Cult. Relics* 9 (47), 3.
- Wang, J. H. (2005). A Study on the Prehistoric Population in the Middle and Lower Reaches of the Yellow River. dissertation/ph.D's thesis. Jinan (Shandong): Shandong University.
- Wang, W. L., Guo, X. N., Kang, N. W., Liu, X. M., and HuChen, K. L. (2015). Archaeology Survey on the Muzhuzhuliang Site in Shenmu County, Shaanxi. *Archaeol. Cult. Relics* 36 (05), 3. doi:10.3969/j.issn.1000-7830.2015.05.001
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., et al. (2016). HaploGrep 2: Mitochondrial Haplogroup Classification in the Era of High-Throughput Sequencing. *Nucleic Acids Res.* 44 (W1), W58–W63. doi:10.1093/nar/gkw233
- Xing, F. L., Li, M., and Sun, Z. Y. (2002). Brief Excavation Report of Xinhua Site, Shenmu, Shaanxi in 1999. *Archaeol. Cult. Relics* 23 (01), 3. doi:10.3969/j.issn.1000-7830.2002.01.001
- Xu, F. (2014). Preliminary Comparison of Shimao and Taosi Archaeology Discovery. *Wenbo* 31 (01), 18. doi:10.3969/j.issn.1000-7954.2014.01.003
- Xu, Y. J. (2004). The Yangshao Culture Survivals' Pedigrees in Later Period in the Loess Plateau. dissertation/ph.D's thesis. Changchun (Jilin): Jilin University.
- Yan, Z. B., and He, N. (2005). Archaeology Survey on the Taosi Site in Xiangfen County, Shanxi. *Acta Archaeol. Sin.* 70 (03), 307.
- Yang, Y. Y. (2016). A Study of the Cultural Development and Landscape Distribution of the Majiayao Culture in Gansu and Qinghai Provinces. dissertation/master's thesis. Lanzhou (Gansu): Lanzhou University.

- Zhang, X. (2015). The Research of Dawenkou Culture. dissertation/ph.D's thesis. Changchun (Jilin): Jilin University.
- Zhang, Z. P. (2004). The Pottery *Li* with Side-Attached Double Lugs of Xinghua Culture. *Palace Mus. J.* 26 (04), 6. doi:10.16319/j.cnki.0452-7402.2004.04.001
- Zhang, Z. P. (1996). Yangshao Age - the Prosperity of Prehistoric Society and the Transition to Civilization. *Palace Mus. J.* 18 (01), 1–44.
- Zhao, J., Liu, F.-E., Lin, S., Liu, Z.-Z., Sun, Z.-Y., Wu, X.-M., et al. (2017). Investigation on Maternal Lineage of a Neolithic Group from Northern Shaanxi Based on Ancient DNA. *Mitochondrial DNA Part A* 28 (5), 732–739. doi:10.1080/24701394.2016.1177039
- Zhou, S. C. (2017). The Origin of Chinese Civilization Yellow River Basin Center Talked about the Theoretical Evolution of Pluralistic Unity Theory - a New Investigation from the Perspective of Academic History. *West. Historiogr.* 1 (00), 11–33.
- Zhou, W. Z. (2015). The Ethnic and Fusion of Northern Shaanxi in Historical Period. *Northwest Ethnol. Ser.* 14 (02), 14.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xue, Wang, Shao, Dai, Sun, Gardner, Chen, Guo, Di, Pei, Wu, Zhang, Cui, Cao, Liu, Dai, Feng, Yang, Ping, Zhang, He and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.