

## GENETICS

## Ancient Xinjiang mitogenomes reveal intense admixture with high genetic diversity

Wenjun Wang<sup>1,2,3\*</sup>, Manyu Ding<sup>2,3,4\*</sup>, Jacob D. Gardner<sup>2,4\*</sup>, Yongqiang Wang<sup>5\*</sup>,  
Bo Miao<sup>1,2,3</sup>, Wu Guo<sup>6</sup>, Xinhua Wu<sup>6</sup>, Qiurong Ruan<sup>5</sup>, Jianjun Yu<sup>5</sup>, Xingjun Hu<sup>5</sup>,  
Bo Wang<sup>7</sup>, Xiaohong Wu<sup>8</sup>, Zihua Tang<sup>9</sup>, Alipujiang Niyazi<sup>5</sup>, Jie Zhang<sup>5</sup>, Xien Chang<sup>5</sup>,  
Yunpeng Tang<sup>10</sup>, Meng Ren<sup>10</sup>, Peng Cao<sup>2</sup>, Feng Liu<sup>2</sup>, Qingyan Dai<sup>2</sup>, Xiaotian Feng<sup>2</sup>,  
Ruowei Yang<sup>2</sup>, Ming Zhang<sup>2,3,4</sup>, Tianyi Wang<sup>2,3,10</sup>, Wanjing Ping<sup>2,3</sup>, Weihong Hou<sup>2,3</sup>,  
Wenyong Li<sup>5</sup>, Jian Ma<sup>10</sup>, Vikas Kumar<sup>2,3†</sup>, Qiaomei Fu<sup>2,3,4†</sup>

Copyright © 2021  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

Xinjiang is a key region in northwestern China, connecting East and West Eurasian populations and cultures for thousands of years. To understand the genetic history of Xinjiang, we sequenced 237 complete ancient human mitochondrial genomes from the Bronze Age through Historical Era (41 archaeological sites). Overall, the Bronze Age Xinjiang populations show high diversity and regional genetic affinities with Steppe and northeastern Asian populations along with a deep ancient Siberian connection for the Tarim Basin Xiaohe individuals. In the Iron Age, in general, Steppe-related and northeastern Asian admixture intensified, with North and East Xinjiang populations showing more affinity with northeastern Asians and South Xinjiang populations showing more affinity with Central Asians. The genetic structure observed in the Historical Era of Xinjiang is similar to that in the Iron Age, demonstrating genetic continuity since the Iron Age with some additional genetic admixture with populations surrounding the Xinjiang region.

## INTRODUCTION

The Xinjiang region in northwestern China has served as a crossroads for East and West Eurasian migrations for thousands of years. As early as the Bronze Age (BA), Xinjiang hosted diverse cultures influenced by ancient Steppe, Siberian, Central Asian, and Northeast Asian populations (1–6). As suggested by many archaeological findings, these cultural influences on Xinjiang varied by region and time period (1–6). Archaeological studies of northern Xinjiang have revealed connections with the Afanasievo (~3300 to 2500 BCE) and Chemurchek (~2750 to 1900 BCE) cultures present in the Altai Mountains (1–3). BA cemeteries in western Xinjiang contain materials associated with mobile transportation and advanced metallurgy, which were likely derived from the Andronovo culture (~1700 to 1500 BCE) in the western Steppe and Tianshan region (7–11). There was a Central West Asian connection with Xinjiang in the BA through the Inner Asian Mountain Corridor (IAMC), which likely introduced agriculturally important crops, such as wheat and barley, and an East Asian connection through the Hexi Corridor, which introduced broomcorn millet in Xinjiang (12–14). In addition, BA populations in eastern Xinjiang share a cultural connection with East Asians from the Gansu and Qinghai (Gan-Qing) region of northern China (3, 13–17). Past studies on two BA sites (Xiaohe and

Tianshanbeilu), using a limited number of Y-chromosomal single-nucleotide polymorphisms (SNPs) and the hypervariable region of mitochondrial DNA (mtDNA), could not resolve the past genetic history of Xinjiang (18–20).

During the Iron Age (IA; ~800 to 200 BCE), nomadic groups from the Eurasian Steppe affected different regions of Xinjiang. One such group was the Scythians who were a confederation of several populations, such as the Tagar, Pazyryk, and Sakas (18, 21). Some burial custom records suggest that the Middle BA southern Siberian Okunevo culture, which had a limited amount of Steppe ancestry, also influenced northern Xinjiang (22–25).

A recent genomic study of a single IA site reported Steppe-related ancestry in eastern Xinjiang (26). Ancient genomic studies in regions around Xinjiang, mostly in the Steppe region, further support the widespread population movement and admixture of western Steppe-related ancestry in the IA (21). However, the extent of Steppe-related ancestry across Xinjiang is unknown without more ancient DNA. Later, after 200 BCE, Xinjiang was dominated by many important nomadic confederations (such as the Xiongnu and Turkic Khaganates) and Han (21, 27). These groups notably influenced this region, and the frequent transitions of power suggest that the Historical Era (HE) was also a culturally mixed period, but whether Xinjiang population structure was affected by these cultural shifts is indeterminable without ancient DNA. Overall, the population genetic structure of ancient Xinjiang remains uncharacterized, as well as the genetic changes from the BA to IA and into the HE. Linguistically, the presence of Tocharian and Khotanese language is also an important question to explore (28). Genomic studies on present-day Xinjiang populations indicate a complex genetic structure with frequent migration and genetic admixture (29, 30). However, with ancient DNA from only a few Xinjiang sites, our ability to reconstruct a complete picture of past population structure and admixture is limited. Therefore, obtaining ancient genetic data from BA, IA, and Historical populations is critical for

<sup>1</sup>College of Life Sciences, Northwest University, Xi'an 710069, China. <sup>2</sup>Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. <sup>3</sup>Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing 100044, China. <sup>4</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>5</sup>Institute of Cultural Relics and Archaeology in Xinjiang, Urumqi 830011, China. <sup>6</sup>Institute of Archaeology, Chinese Academy of Social Sciences, Beijing 100710, China. <sup>7</sup>Xinjiang Uygur Autonomous Region Museum, Urumqi 830002, China. <sup>8</sup>School of Archaeology and Museology, Peking University, Beijing 100871, China. <sup>9</sup>Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China. <sup>10</sup>School of Cultural Heritage, Northwest University, Xi'an 710069, China.

\*These authors contributed equally to this work.

†Corresponding author. Email: fuqiaomei@ivpp.ac.cn (Q.F.); vikas@ivpp.ac.cn (V.K.)

characterizing the spatiotemporal changes in Xinjiang's population structure.

## RESULTS

### Overview of data

We captured mtDNA from 237 ancient individuals from 41 archaeological sites across Xinjiang (listed in table S1), with samples dating to ~4962 to 500 years before the present (BP). Of these 237 mtDNA samples, 235 of them have low contamination rates ( $\leq 4\%$ ), while for the other two samples with higher contamination rates ( $> 4\%$ ), we only used fragments with damaged signatures characteristic of ancient DNA to avoid contamination. In total, we obtained complete mtDNA genomes from 237 individuals sequenced to between 31 and 5515 $\times$  coverage (table S1).

For the BA, we collected 63 samples spanning the time period of ~4962 to 2900 BP (table S1), including 24 individuals from North Xinjiang in the Early and Middle BA (EMBA; ~4800 to 4000 BP); among them, 18 individuals were associated with the Chemurchek culture (NChemur\_EMBA, 4811 to 3965 BP), and two were associated with the Afanasievo culture (NAfana\_EMBA; 4570 to 4426 BP). We obtained three samples associated with the Afanasievo culture from the Ili Valley in western Xinjiang (WAfana\_EMBA; ~4962 to 4840 BP) and combined them with samples associated with the Afanasievo from North Xinjiang to form the NWAfana\_EMBA group. The other sites, Songshugou (NSSG\_EMBA;  $n = 3$ , 4237 to 4087 BP) and Habahe ( $n = 1$ , ~4500 BP), with no known information on their archaeological cultures were analyzed separately. From eastern Xinjiang, we collected one individual from the BA Nanwan site (E\_BA; ~3600 to 3000 BP) and an additional 25 samples dating to the Late BA (E\_LBA; ~3000 to 2900 BP). We merged the haplogroup information of our single Nanwan individual with 29 previously published Tianshanbeilu individuals to comprise the E\_BA group, given the close archaeological relationship between the two sites (18). In addition, we collected 10 samples from the fourth to fifth layers of the Xiaohe archaeological site in the eastern Tarim Basin, southeastern Xinjiang (SEXiaohe\_BA; ~3929 to 3572 BP) (figs. S1 and S2) (20). These groupings represent the eastern, western, northern, and southeastern Xinjiang populations during the BA, as indicated by the first letters of their labels (Fig. 1).

For the IA, we collected 128 samples (~2900 to 2000 BP) across Xinjiang (Fig. 1), of which 27 were from eastern Xinjiang (E\_IA), 15 from northern Xinjiang (N\_IA), 55 from the Ili region in western Xinjiang (W\_IA), and 31 from southern Xinjiang (S\_IA). We did not merge the different S\_IA archaeological sites into one large group because of their high cultural heterogeneity and wide geographic distribution (31–36); we analyzed them as separate groups. The S\_IA groups SZGLK\_IA ( $n = 19$ ) and SWJEZKL\_IA ( $n = 12$ ) were from the Tarim Basin, while the SWJEZKL\_IA was from the high-altitude region of southwestern Xinjiang, adjacent to the Pamir Plateau (Fig. 1). We also collected and analyzed the haplogroup information of 10 published IA individuals (~2200 BP) from the Shirezigou site of eastern Tianshan (grouped as SRZG\_IA) (26).

In addition to the BA and IA mitogenomes, we also sequenced mtDNA from 46 HE (~2000 to 500 BP) individuals. Considering the sample size and highly mixed archaeological cultures of HE sites, we first grouped the samples based on their geographical locations and then additionally subdivided geographical groups based on their archaeological cultures. There are five groups in total, one

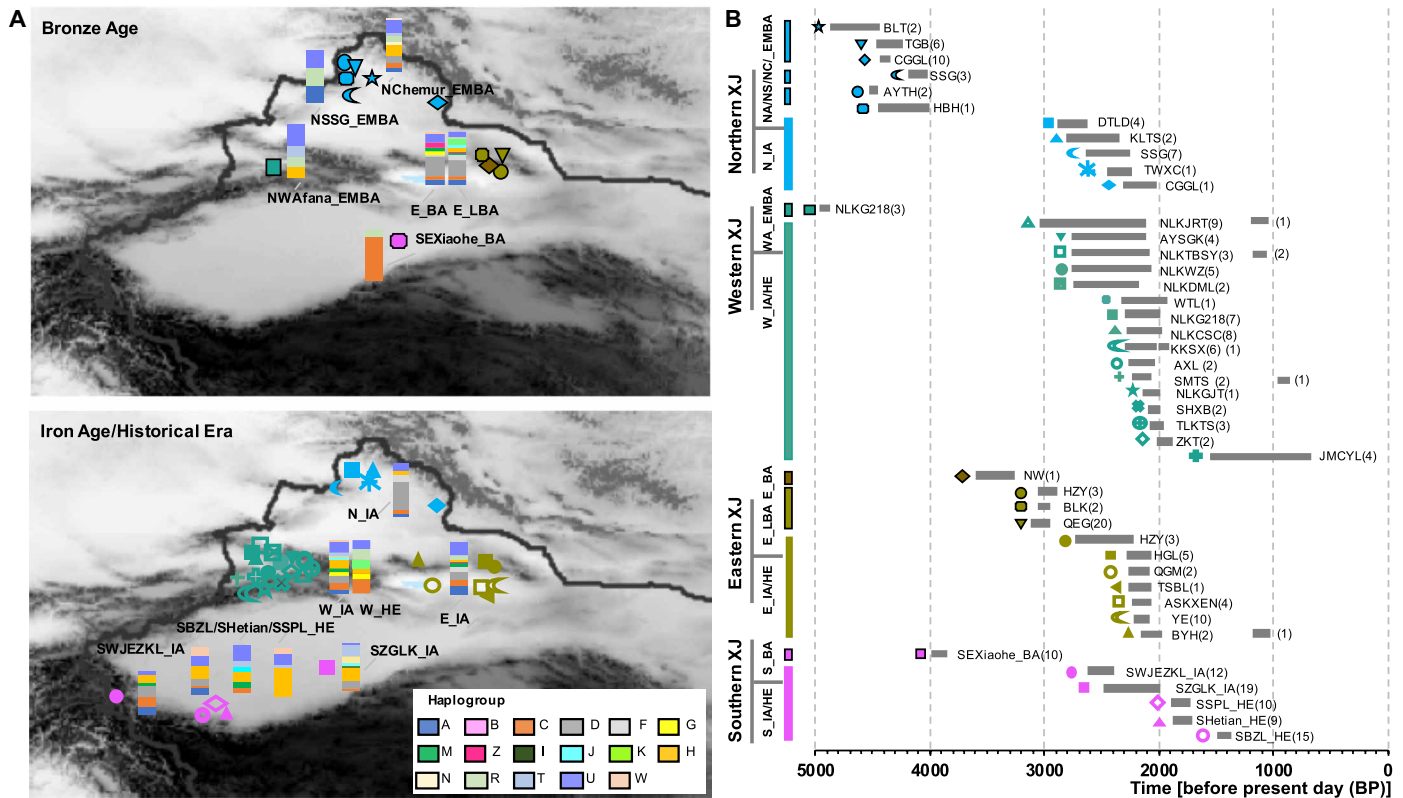
from western Xinjiang (W\_HE;  $n = 11$ ), three from southern Xinjiang (SBZL\_HE,  $n = 15$ ; SSPL\_HE,  $n = 10$ ; and SHetian\_HE,  $n = 9$ ), and one individual from the Baiyanghe site in eastern Xinjiang (E\_HE). In addition, we obtained the published ancient ( $n = 738$ ) and present-day ( $n = 7085$ ) mtDNA data from regions surrounding Xinjiang and grouped all these populations into several subgroups based on previous genetic studies (see Box 1 and table S1 for additional details).

In general, for all the sampled ancient Xinjiang populations, we find a positive and significant correlation coefficient between maternal genetic distance ( $F_{ST}$ ) and geographic distance ( $R^2 \approx 0.0327$ ,  $P \approx 0.018$ ) (fig. S3), using the Mantel test (37, 38). Thus, ancient Xinjiang populations were likely highly admixed and had low geographic structuring. The genetic comparison of BA, IA, and HE Xinjiang populations also revealed variation in nucleotide diversity ( $\pi$ ) (fig. S4). IA and HE populations, generally, showed higher  $\pi$  compared to their BA populations, indicative of elevated population migration and admixture during the IA and HE relative to the BA. Among the IA and HE populations, W\_HE ( $0.0024 \pm 0.0013$ ) showed the highest diversity, with the lowest diversity observed in southern Xinjiang populations ( $0.0015 \pm 0.0008$  to  $0.0020 \pm 0.0010$ ) (fig. S4).

### Genetic origins and complexity of BA Xinjiang populations

To determine the genetic differences and affinities among the BA Xinjiang groups, we first conducted a discriminant analysis of principal components (DAPC) based on haplogroup frequency. PC1 explains population variation from east to west geographically, and PC2 explains the variation from north to south (Fig. 2A and fig. S5). In general, all the populations were divided into four main clusters: northeastern Asian (NEA: Siberian and northern East Asian), southeastern Asian (SEA: southern East Asian and Southeast Asian), central Steppe, and European (Turan and European). All the ancient Xinjiang samples lie on a cline extending from the NEA populations to the central Steppe and European clusters (Fig. 2A and fig. S5), suggesting that these ancient Xinjiang populations had varying degrees of relatedness to NEA, central Steppe, and European populations.

We find that the EMBA Xinjiang individuals from northern and western Xinjiang, associated with the Afanasievo culture (NWAfana\_EMBA), are surrounded by western Steppe-related populations (WSteppe\_EMBA and WSteppe\_MLBA) (Fig. 2A and fig. S5). In contrast, the EMBA individuals associated with the Chemurchek culture (NChemur\_EMBA) and individuals from the Songshugou site in northern Xinjiang (NSSG\_EMBA) form their own separate cluster surrounded by other central Steppe populations (Fig. 2A and fig. S5). High proportions of U, H, and R haplogroups are observed, which were reported primarily in BA Steppe populations (table S1). Although there is no significant genetic differentiation among these EMBA individuals ( $F_{ST} \approx 0.00$ ,  $P > 0.05$ ) (Fig. 2B, fig. S6, and table S2), we find that only NChemur\_EMBA shows significant genetic differentiation from both western Steppe populations (WSteppe\_EMBA:  $F_{ST} \approx 0.045$ ,  $P \approx 0.005$ ; WSteppe\_MLBA:  $F_{ST} \approx 0.042$ ,  $P \approx 0.006$ ) (Fig. 2B and table S2). NChemur\_EMBA also shows significant genetic differentiation from CSteppe\_MLBA ( $F_{ST} \approx 0.057$ ,  $P \approx 0.002$ ) but not from CSteppe\_EMBA ( $F_{ST} \approx 0.029$ ,  $P \approx 0.16$ ) (Fig. 2B and table S2), consistent with its position on the DAPC plot (fig. S5C). Moreover, in the median-joining networks, the WSteppe\_EMBA individuals cluster with NWAfana\_EMBA in haplogroups U4, U5, and H15 (Fig. 3D and fig. S7B); with NChemur\_EMBA in



**Fig. 1. Geographic locations and haplogroups of ancient Xinjiang samples. (A)** Geographic locations and haplogroups of BA and IA/HE samples from eastern (E), western (W), southern (S), and northern (N) Xinjiang (XJ). The different colored and shaped symbols overlaying Xinjiang represent different sites (group names in the map are consistent with those in the manuscript). We merged haplogroup information of our one Nanwan (NW) individual with 29 previously published Tianshanbeilu (TSBL) individuals to comprise the E\_BA group. Colored bar plots represent the frequencies of various haplogroups, shown on the bottom right. **(B)** Timeline of Xinjiang sites with the number of individuals in parentheses. NS, NC, and NWA\_EMBA are the abbreviations of NSSG\_EMBA, NChemur\_EMBA, and NWAfana\_EMBA, respectively.

U4, U5, H2, H6a, and W3 (Fig. 3D and fig. S7, B and D); and with NSSG\_EMBA in U4 (Fig. 3D and fig. S7B) (39, 40). The two haplogroups H2 and H5 are present in the western Steppe-related populations, and H6a appears in the populations related to the Okunevo culture present in the Altai region (fig. S7B and table S1) (40). We also find an NEA connection (Baikal\_EBA) with NChemur\_EMBA supported by the appearance of haplogroup D4j (table S1), which differed in these two populations by only four mutations in the network analysis and appeared in northern East Asia, including the southern Siberian region (Fig. 3E and fig. S7A) (41). The one EMBA sample from HBH has the haplogroup U5 (table S1), suggesting a more western Steppe-related connection. Therefore, we demonstrate that both western and northern Xinjiang populations have considerable western Steppe-related ancestry during the EMBA.

We find that eastern groups (E\_BA and E\_LBA) cluster separately from the EMBA individuals in northern and western Xinjiang. Both of the eastern groups cluster with ancient and present-day NEA in the DAPC (Fig. 2A and fig. S5). E\_BA and E\_LBA harbor high proportions of the haplogroup D (~36.70 and 32.00%, respectively), which is a common lineage in ancient and present-day NEA populations (42, 43) including northern Chinese (18.20 to 44.80%) and ancient Mongolians (31.20%) (Figs. 1A and 3E and table S3). E\_LBA also shows nonsignificant genetic distances to some of these NEA populations, specifically two ancient Gan-Qing populations (GQQijia\_BA and GQKayue\_LBA;  $F_{ST} < 0.05$ ,  $P > 0.05$ ) and four

present-day populations (Japanese, Mongolian, Tu, and Oroqen;  $F_{ST} < 0.03$ ,  $P > 0.05$ ) (fig. S6, B and C, and table S2). Although both E\_BA and E\_LBA have the western Steppe-related haplogroup U, they show a higher proportion of lineages from NEA than from Europe, with more European lineages appearing in later samples (20% in E\_BA and 36% in E\_LBA) (table S3). This pattern is consistent with DAPC in which E\_LBA plots closer to West Eurasians compared to E\_BA (Fig. 2A). Moreover, haplogroup D4b2b4 is found in both the Xiongnu and E\_LBA (Fig. 3E), which suggests a direct relationship between E\_LBA and Xiongnu populations due to the presence of shared NEA ancestry. Thus, E\_BA and E\_LBA populations show more NEA connections, but the presence of western Steppe-related lineages (U, 16.7% in E\_BA and 8% in E\_LBA) also supports additional connections to the western Steppe-related populations (table S3).

Although SEXiaohe\_BA clusters into the NEA groups in DAPC, which is similar to E\_BA and E\_LBA, they show more affinity for populations with ancient and present-day Siberian ancestry (Fig. 2A and fig. S5). SEXiaohe\_BA has a high proportion of the C4 haplogroup (six of seven individuals) present in ancient and present-day Siberian populations, including NEA and Shamanka populations from near the Lake Baikal region of South Siberia (Fig. 3A and fig. S7A). This population is unique in yielding significant genetic distances compared to all other ancient and present-day populations ( $F_{ST} > 0.11$ , most  $P$  values  $< 0.04$ ), including other BA Xinjiang groups, but it has the lowest genetic distances with three present-day

**Box 1. Group composition and new findings in this study.****Ancient groups for published individuals**

> **European populations (Euro):** Baltic\_BA, BellBeaker\_LN, Hungary\_BA, Unetic\_LNBA, Scandi\_LNBA; early European farmer: LBKH\_EN, Iberia\_EN, Iberia\_Chal, and Anatolia\_N; western Steppe: WSteppe\_EMBA (Yamnaya\_EBA, Afanasievo\_EBA, Poltavka\_BA, and Potapovka\_BA), WSteppe\_MLBA (CordedWare\_BA, Sintashta\_MLBA, Srubnaya\_MLBA, Andronovo\_MLBA, and WKazakh\_MLBA); Anatolia\_BA.

> **Turan populations:** Turkmen\_EN, Turkmen\_MLBA, Uzbek\_MLBA, Iran\_C, Iran\_EMBA.

> **Bronze Age central Steppe populations (CS\_BA):** Okunevo\_EMBA, CSteppe\_EMBA (Kazakh\_EBA), CSteppe\_MLBA (Kazakh\_MLBA and Krasn\_MLBA), Karasuk\_MLBA.

> **Iron Age central Steppe populations (CS\_IA):** Sarmatian\_IA, CenSaka\_IA, TianSaka\_IA, TianHun\_HE, Nomad\_IA, Nomad\_Med, WuKang\_HE, Tagar\_IA.

> **Northeastern Asian populations (NEA):** Siberian: Baikal\_EN (Shamanka\_EN), Baikal\_EBA (Shamanka\_EBA and Ustlda\_EBA); northern East Asian: Khovsgol\_LBA, GQMajiaY\_EBA, GQQijia\_BA, GQKayue\_LBA, LTP\_IA, Xiongnu\_HP.

> **Southeastern Asian populations (SEA):** HTP\_IA.

**Present-day groups for published individuals**

> **Present-day northeastern Asian populations (pdNEA):** Siberian: Yakut, Even, Evenk, Koryak, Yukaghir, Chukchi; northern East Asia: Japanese, Korean, Barghut, Daur, Mongolian, Tu, Oroqen, Hezhen, Xibo, NHan.

> **Present-day southeastern Asian populations (pdSEA):** Southern East Asian: Dai, Miao, She, Yi, Lahu, Naxi, Tujia, SHan, TibetC, TaiwanC; Southeast Asia: Cambodia, India, Laos, Myanmar, Nepal, Philippines, Thailand, Vietnam.

> **Present-day populations in and around Xinjiang (pdCA/XJ):** Buryat, AltaianSib, KazSib, UygurXJ, KyrgyzXJ, SarTajikXJ, WakTajikXJ, LowTajik, PamirTajik, KazTajik, Pakistan.

> **Present-day Europeans and West Asian populations (pdEurWA):** Caucasus, European, WestAsia.

**Ancient groups for samples in four regions of Xinjiang****> Northern and western Xinjiang**

NChemur\_EMBA: EMBA individuals ( $n=18$ , ~4811–3965 BP) associated with Chemurcheke culture from northern Xinjiang show a connection with western Steppe, central Steppe, and NEA populations.

NSSG\_EMBA: EMBA individuals ( $n=3$ , ~4237–4087 BP), with unknown culture, show a connection with western Steppe and NEA populations.

NWafana\_EMBA: EMBA individuals ( $n=5$ , ~4962–4426 BP) associated with Afanasievo culture from northern and western Xinjiang show an affinity with western Steppe populations.

N\_IA: IA individuals ( $n=15$ , ~2900–2000 BP) from northern Xinjiang are closely related to NEA.

W\_IA: IA individuals ( $n=55$ , ~3000–2000 BP) from western Xinjiang are closely related to both NEA and European lineages and share lineages with Turan.

W\_HE: HE samples ( $n=11$ , ~2000–500 BP) from western Xinjiang are closely related to both NEA and European lineages and share lineages with Turan.

**> Eastern Xinjiang**

E\_BA: One BA sample (~3600–3000 BP) from eastern Xinjiang.

E\_LBA: 25 LBA samples (~3000–2900 BP) from eastern Xinjiang show high affinity with NEA.

E\_IA: 27 IA individuals (~2800–2000 BP) from eastern Xinjiang share ancestry with both NEA and European lineages but have more affinity with NEA.

E\_HE: One individual (~1200–1000 BP) from Historical Era.

**> Southern Xinjiang in Tarim Basin**

SEXiaohe\_BA: 10 BA samples from the 4th to 5th layers of the Xiaohe site in southeastern Xinjiang (~3929–3572 BP) are closely related to NEA (Siberian) populations.

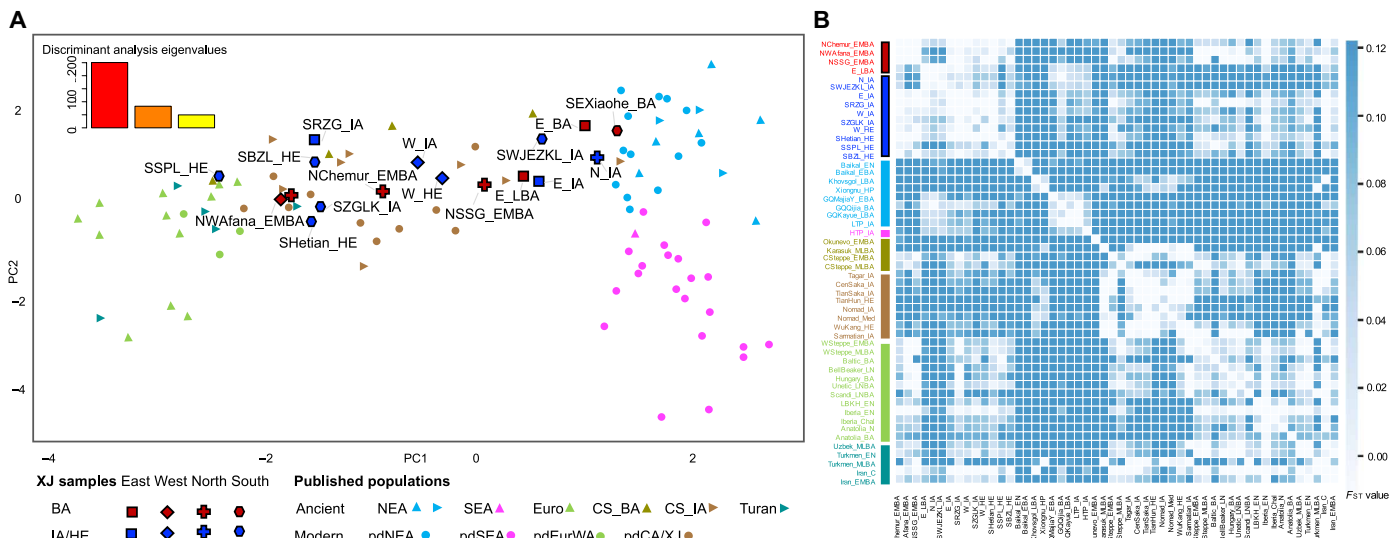
S\_IA: Two IA sites, SZGLK\_IA (~2488–1968 BP,  $n=19$ ) and SWJEZKL\_IA (~2680–2348 BP,  $n=12$ ). SZGLK\_IA has an affinity with MLBA individuals in western Steppe and Turan; SWJEZKL\_IA, from southwestern Xinjiang, near the Pamir Plateau at a high-elevation region, shows a more NEA connection. S\_HE: Three HE sites, SSPL\_HE (~1876–1724 BP,  $n=10$ ), SHetian\_HE (~1812–1605 BP,  $n=9$ ), and SBZL\_HE (~1529–1412 BP,  $n=15$ ), show an affinity with western Steppe and Turan populations.

populations from Siberia (Even, Evenk, and Yakut:  $F_{ST} < 0.13$ ) (table S2). These results are consistent with previous studies on Xiaohe (19, 20).

We also find the mtDNA haplogroup R1b in BA Xinjiang samples (NChemur\_EMBA,  $n = 2$ ; NWafana\_EMBA,  $n = 1$ ; NSSG\_EMBA,  $n = 1$ ; SEXiaohe\_BA,  $n = 1$ ) and in IA and HE populations from eastern and western Xinjiang (E\_IA,  $n = 1$ ; W\_IA,  $n = 1$ ; W\_HE,  $n = 2$ ) (table S1), which was reported not only in East European Hunter-Gatherers (Karelia) (44) but also in Botai (40) and Dali (28) individuals from Kazakhstan. Moreover, the haplogroup K1b2 was shared among the Botai (40) and western Steppe-related populations as well as our LBA samples from eastern Xinjiang (E\_LBA) (table S1). The R1b median-joining network shows that the EMBA sample (3012 to 2890 cal BCE) from northern and western Xinjiang, associated with the Afanasievo (NWafana\_EMBA), plots in the center of the network and was separated from Botai by only a single mutation (Fig. 3B). This branch, in turn, is associated with NSSG\_EMBA and another branch that includes an individual from the Dali site (Fig. 3B). This may suggest either a deep ancestry connection with an Ancient North Eurasian (ANE) population or some genetic connections with geographically proximal populations from Kazakhstan (Dali and Botai) (28, 40). We also find the R1b haplogroup in one of the individuals from the Xiaohe population (Fig. 3B), which may also suggest a North Xinjiang connection with Xiaohe people (11). Thus, during the BA, the northwestern Xinjiang populations showed a high genetic affinity for western Steppe-related cultures, such as the Afanasievo and Chemurcheke, and the southeastern Xinjiang populations for NEA and South Siberian populations (Fig. 4A), suggesting a scenario of complex interactions with the neighboring populations and communities of diverse cultural backgrounds.

**Greater cultural and haplogroup diversity in IA Xinjiang**

To better understand the population movements and changes in Xinjiang during the IA, we analyzed 128 samples across Xinjiang (Fig. 1). In general, all the Xinjiang IA groups except N\_IA show higher genetic haplogroup diversity than those in the BA (Fig. 1A), suggesting more migrations and communication from East and West Eurasia. The genetic differentiation ( $F_{ST}$ ) between the IA groups is mostly nonsignificant ( $P > 0.05$ ), suggesting high levels of admixture in the IA (Fig. 2B and table S2). In the DAPC analysis (Fig. 2A and fig. S5C), we find that the N\_IA clusters close to NEA populations (such as ancient GQMajiaY\_EBA and Khovsgol\_LBA in northern China, present-day Xibo in northern China, and present-day Yakut in Siberia), which is contrary to the EMBA northern Xinjiang populations who cluster close to those with western Steppe-related ancestry. This high affinity for NEA is also represented by two major EA haplogroups, D (53.30%) and F (13.30%) (Fig. 1A and table S3), and low  $F_{ST}$  values with North China populations, such as GQMajiaY\_EBA, GQQijia\_BA, GQKayue\_LBA, and LTP\_IA (table S2). The



**Fig. 2. The genetic results for the ancient Xinjiang samples and other ancient and present-day Eurasians. (A)** DAPC based on haplogroup frequency; the eigenvalues of the first and second PCs are shown on the top right. The different colors on the plot represent different groups made under unsupervised classification. The shapes with black frames represent the ancient Xinjiang samples with the BA samples in dark red and the IA/HE samples in dark blue. The populations plotted as triangles are ancient populations, and the circles are present-day humans. NEA, northern East Asian (upward pointing triangles), and Siberia (triangles pointing to the right); SEA, southeastern Asia; Euro, Europe; CS, central Steppe populations in BA (CS\_BA; upward pointing triangles) and IA (CS\_IA; triangles pointing to the right); pdCA/XJ, present-day populations in and around Xinjiang. **(B)** The genetic distance ( $F_{ST}$ ) heatmap plot of ancient Xinjiang samples and ancient Eurasians. The different labels represent genetically distinctive groups corresponding to those in the DAPC. Values with  $F_{ST} \approx 0.00$  are in white, representing a close genetic relationship. SEXiaoho\_BA was removed from the  $F_{ST}$  heatmap, considering the significantly large genetic distances ( $F_{ST} > 0.10$ ) between them and other groups.

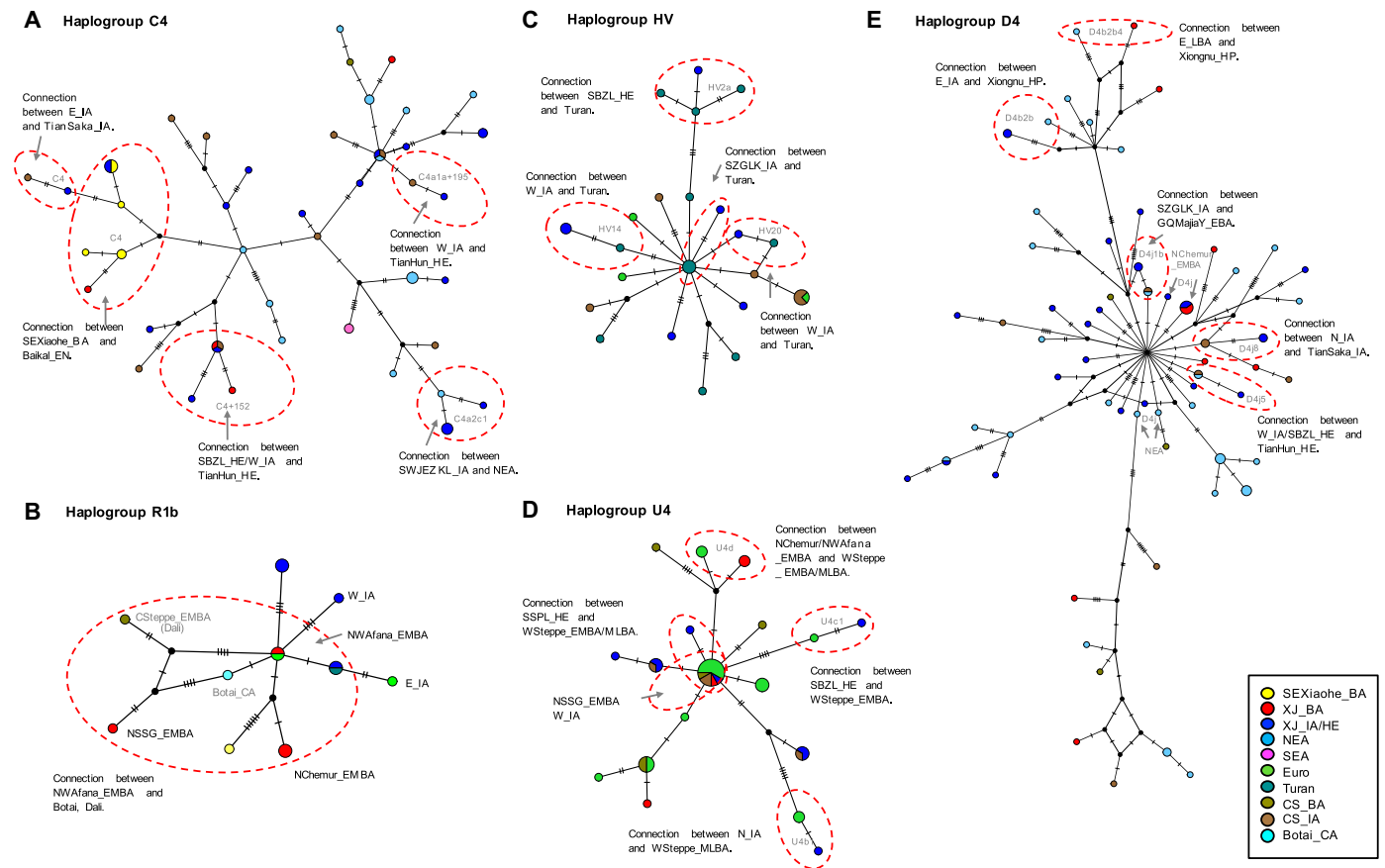
median-joining networks further suggest a connection between N\_IA and NEA individuals (haplogroups D4e and D5a), along with some Saka individuals (TianSaka\_IA, D4j8) (Fig. 3E and fig. S7E). In addition, we also observe western Steppe-related haplogroups U4 (6.70%) and H5 (6.7%) and the Turan-related haplogroup U7 (6.7%) (Fig. 3D and fig. S7, B and C) (45, 46), suggesting a genetic connection with these populations.

There are two groups of populations that include Xinjiang E\_IA and the published samples from SRZG\_IA. We observe that both E\_IA and SRZG\_IA have very low genetic distances with GQQijia\_BA from northern China ( $F_{ST} < 0.05$ ,  $P > 0.05$ ), consistent with their BA samples, E\_LBA (Fig. 2B and table S2). In DAPC, although both E\_IA and SRZG\_IA plot in the central Steppe cluster, E\_IA shows a slightly greater affinity to the NEA populations, whereas the published samples from SRZG\_IA group closer to those with western Steppe-related ancestry (Fig. 2A and fig. S5). We also observe higher proportions of the western steppe-related U haplogroup in SRZG\_IA (40%) compared to the E\_IA (22.20%), supporting the DAPC plot (Fig. 1A and table S3). In addition, we also find the NEA haplogroup D (14.8%) in E\_IA, represented by D4b2b (in the Xiongnu) (Fig. 3E, fig. S7A, and table S1). We also observe Central Asian haplogroups. The Turan-specific haplogroup T2d1 (Uzbek\_MLBA) in E\_IA likely reflects additional affinities shared with Turan populations (fig. S7C). Therefore, E\_IA populations not only show both NEA and European haplogroups (Fig. 1 and table S3) but also share additional genetic affinity with Turan populations.

In DAPC, the W\_IA populations lie on the cline of NEA and European populations and group close to the IA populations in the central Steppe (e.g., Saka and Hun) (Fig. 2A and fig. S5, A and C). This East and West Eurasian affinity could be seen in the diverse array of East and West Eurasian haplogroups compared to the other three IA geographical regions (Fig. 1A). The W\_IA includes the two

major European haplogroups U (20.40%) and H (18.50%) and the NEA haplogroups C (14.80%) and D (11.10%) (Fig. 1A and table S3). The connection to the WSteppe\_MLBA lineages is also observed with sub-haplogroups T2b34 and U5a2a1 but not in the BA samples from the same region (fig. S7B and table S1). The median-joining network of T2b34 also shows a W\_IA and WSteppe\_MLBA connection (fig. S7B). This high haplogroup diversity is further reflected in the  $F_{ST}$  values whereby many of the NEA and European populations show very low  $F_{ST}$  values with W\_IA (table S2). In addition, the presence of the Turan-specific haplogroup HV (HV18 and HV20) and W (W3b) (Fig. 3C and fig. S7, C and D) indicates the possible migration of people from the Turan through IAMC to the Ili region. We also observe some important outliers in W\_IA, such as the presence of haplogroups G3a3 (Xiongnu\_HP), C4a1a+195 and C4+152 (TianHun\_HE), and H101 (CenSaka\_IA), suggesting a possible connection with Steppe nomadic groups (table S1). Also, the phylogenetic networks of C4, G3a3, and H101 show direct connections with few differences between the W\_IA and the IA Steppe nomadic Saka, Xiongnu, and Hun populations (Fig. 3A and fig. S7D), suggesting the potential role of the Ili region in the migration of the Saka, Xiongnu, and Hun in ancient times.

Last, the SZGLK\_IA from South Xinjiang shows a predominant Steppe and Central Asian connection as they lie close to the Steppe and Central Asian populations in DAPC, whereas SWJEZKL\_IA lies close to populations with NEA and central Steppe ancestry (Fig. 2A and fig. S5) with high frequencies of NEA haplogroups C (25.00%) and D (25.00%) (Fig. 1A and table S3). Haplogroup analysis shows that SZGLK\_IA has a relatively higher frequency of H (26.3%) and U (5.3%) haplogroups (table S3), which are present in ancient West Eurasian populations, such as WSteppe\_MLBA (fig. S7B); SZGLK\_IA also has low  $F_{ST}$  values ( $< 0.05$ ,  $P = 0.005$ ) with WSteppe\_MLBA (table S2). The connection to the Steppe\_MLBA is



**Fig. 3. Median-joining haplogroup networks.** The median-joining network of the haplogroups C4 (A), R1b (B), HV (C), U4 (D), and D4 (E) related to ancient northeastern Asian (NEA), Botai/Dali, European (Euro), Turan, and BA and IA populations from the central Steppe (CS\_BA and CS\_IA). The Euro group consists mainly of western Steppe-related individuals, and the CS\_IA group contains the Saka, Hun, and Nomad populations. The size of the circles represents the proportion of each haplotype. The lengths of lines between nodes represent the number of mutations between two haplotypes. The different population groups are shown by different colors that are consistent with those groups in DAPC and  $F_{ST}$  heatmap. Some ancient Xinjiang samples are annotated with black labels.

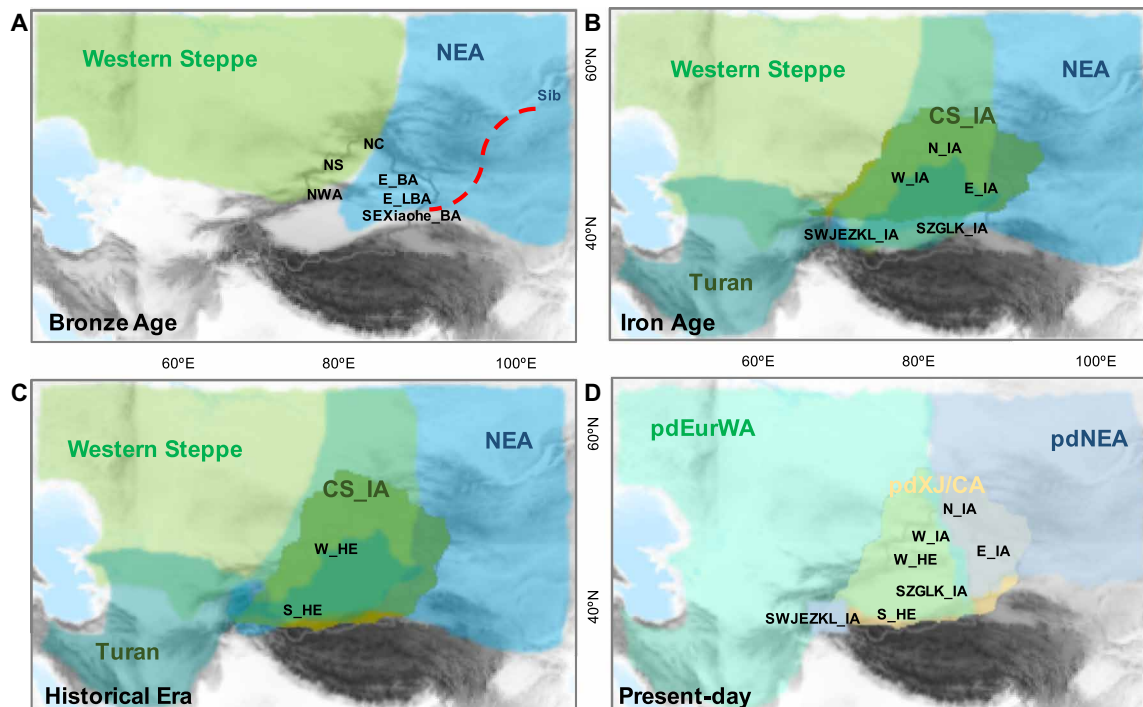
further augmented by the presence of haplogroups T2b34, H5a1, U5a2a1, and N1a1a1a, present in WSteppe\_MLBA populations but not in WSteppe\_EMBA (table S1 and fig. S7B). Moreover, the haplogroup N1a1a1a has a high frequency (about 6 of 73) in early European farmers (LBK\_EN and Anatolia\_N) but not in European Hunter-Gatherers (table S1 and fig. S7B). The presence of the HV12 and R2+13500 haplogroups in SZGLK\_IA also reveals a Central Asian connection from Turan (Fig. 3C and fig. S7C). A close affinity with Central Asians was further found with lower  $F_{ST}$  values comparing SZGLK\_IA with Iran\_C ( $=0.00$ ) and Turkmen\_EN ( $\approx 0.01$ ) (Fig. 2B and table S2). In addition, the SWJEZKL\_IA too had a Turan-related lineage (H13a2a), indicating some connections with Turan (table S1 and fig. S7D).

Overall, the IA populations show a distinct population structure across Xinjiang despite intense admixture. The N\_IA and high-elevation SWJEZKL\_IA populations show more NEA ancestry, whereas the W\_IA and E\_IA populations show both NEA and European ancestries, and SZGLK\_IA shows more connections with the western Steppe-related populations in the MLBA (Fig. 4B). In addition, all the IA populations show a genetic connection with Turan populations; among these IA samples, E\_IA, W\_IA, and SZGLK\_IA shared more affinity with the Turan populations compared to N\_IA (table S2). This further suggests the important role

played by the IAMC, which probably led to the increased migration of Turan people with ancestries similar to the Bactria Margiana Archaeological Complex culture prevalent in this region (11, 47).

**Genetic continuity between the IA and HE in Xinjiang**

In addition to the BA and IA individuals, to further assess the genetic diversity after the IA time period, we sequenced 46 samples from the HE with a range of archaeological dates between ~2000 and 500 BP. In general, we find that the HE individuals show the lowest  $F_{ST}$  values with ancient Xinjiang populations compared to other ancient populations outside of Xinjiang (table S2). In DAPC, the W\_HE clusters close to other IA individuals of W\_IA, SRZG\_IA, SBZL\_HE, and SZGLK\_IA along with ancient central Steppe individuals (Fig. 2A and fig. S5). The W\_HE populations also show the same NEA lineages (C, 27.30%; G, 9.10%) and European haplogroups (H, 9.10%; U, 18.20%) present in the IA (W\_IA) individuals (Fig. 1A and table S3). They additionally have the Hun (C4a1a+195), Saka (H101), and Turan mtDNA lineages that were also found in the W\_IA (table S1). The close affinity between the W\_HE and the W\_IA is further shown by low  $F_{ST}$  values ( $F_{ST} \approx 0.002$ ,  $P > 0.05$ ) (table S2). The one historical sample from East Xinjiang has the haplogroup D4 (table S1), suggesting a more NEA connection.



**Fig. 4. Inferred maternal population movements around Xinjiang.** Depictions of the main population movements in the BA (A), IA (B), HE (C), and present day (D). The dashed line in (A) represents the connection between SEXiaohe\_BA and ancient Siberian (Sib) populations. The shades of green, blue, brown, and turquoise in (A) to (C) represent the ancient European-, NEA-, central Steppe-, and Turan-specific ancestries, respectively. The shades of light green, blue, and yellow represent lineages found in present-day Europeans and West Asians, NEA, and populations in and around Xinjiang (pdCA/XJ). NS, NC, and NWA in (A) are the abbreviations of NSSG\_EMBA, NChemur\_EMBA, and NWAfana\_EMBA, respectively.

The South HE (S\_HE) individuals from the three sites (SBZL\_HE, SSPL\_HE, and SHetian\_HE) yielded varied affinities. SBZL\_HE is unique among the S\_HE in having haplogroup D (13.3%), while SSPL\_HE contains a high proportion of the western Steppe-related haplogroup H (60%) (Fig. 1A and table S3). Haplogroup frequencies show the same pattern, whereby the SBZL\_HE individuals show a relatively higher NEA affinity (33.3%, haplogroups A, C, and D) compared to the SHetian\_HE (C, 11.1%) and SSPL\_HE (0.00%) individuals (Fig. 1A and table S3), which show higher affinities for western Steppe-related populations (60% H, 20% U, 10% T, and 10% W in SSPL\_HE) (Fig. 1A and table S3), and group close to the western Steppe-related populations in DAPC (Fig. 2A and fig. S5). The three HE populations show low  $F_{ST}$  values with western Steppe-related and Turan populations ( $F_{ST} < 0.05$ ) (table S2), similar to the IA population from the same region (SZGLK\_IA). However, we find the Hun-related haplogroup (D4j5) in some of the S\_HE samples (SBZL\_HE) and in W\_IA (Fig. 3E). The D4 haplogroup network shows that W\_IA and SBZL\_HE are separated by fewer mutations with TianHun\_HE (Fig. 3E), which may reflect a connection between IA western Xinjiang and HE southern Xinjiang, and a southward movement of the Hun from the Ili region into southern Xinjiang.

### Mitogenomic comparisons of BA, IA, and HE Xinjiang individuals

Above all, haplogroup comparisons between different Xinjiang geographical regions across the BA and IA suggest the occurrence of multiple migrations and admixture events (Fig. 1A). The EMBA

populations in northern Xinjiang show higher affinity with western Steppe-related populations, suggested by haplogroups U and H, whereas the IA population (N\_IA) shares more connections with East Eurasians, especially those found in NEA with high proportions of haplogroup D4 (such as D4c1b1, D4e1, and D4o2a that are specific in IA/HE) (Fig. 2A and table S1). We found consistent results in our  $F_{ST}$  analysis where EMBA in northern Xinjiang populations shows a small genetic distance with the western Steppe-related populations ( $F_{ST} = 0.00$ ,  $P > 0.05$  in NWAfana\_EMBA and NSSG\_EMBA) and the N\_IA with the ancient NEA ( $F_{ST} = 0.00$ ,  $P > 0.05$  in GQQijia\_BA) (Fig. 2B and table S2). This shift toward more NEA-related ancestry from the BA to IA period suggests more frequent migrations and admixture between NEA and North Xinjiang populations.

The western Xinjiang populations show fairly consistent haplogroup compositions across the BA and IA periods, with the presence of western Steppe-related haplogroups (U5 and H15), while W\_IA shared some NEA haplogroups (C4, G2, and D4) (table S1). W\_IA plotted between NEA and European populations and clustered with central Steppe populations in DAPC, which was distinct from the EMBA individuals (NWAfana\_EMBA) associated with the Afanasievo (Fig. 2A and fig. S5). W\_IA also shows a Turan connection ( $F_{ST} < 0.03$ , most  $P > 0.05$  in Turkmen\_EN, Iran\_C, and Iran\_EMBA) (table S2). The genetic comparison from the BA to IA thus supports the important role played by the Ili region, where populations with Steppe-related and NEA ancestries thrived for a very long time.

In eastern Xinjiang, both the BA and IA groups show more affinity to NEA populations, but we also find haplogroups from

European populations (Fig. 1A and table S3), further suggesting the presence of NEA and European-admixed populations in this region. The South IA population shows a Turan connection (SZGLK\_IA,  $F_{ST} < 0.03$ ,  $P > 0.05$  in Iran\_C and Iran\_EMBA), similar to the W\_IA (Fig. 2B and table S2), which may reflect the migration of populations from the West (Ili) to South Xinjiang. These highly admixed ancestries in the IA populations continued into the HE. The Historical period in Xinjiang saw varied migrations and settlements of both NEA and European populations, thus reflecting a civilization founded and settled by both East and West Eurasian populations.

To test how different classifications would affect the variance among different Xinjiang groups, we conducted analysis of molecular variance (AMOVA) tests among the ancient Xinjiang populations (table S4). When compared to other groups, the significantly highest  $F_{CT}$  value (variance among groups;  $F_{CT} = 0.054$ ,  $P = 0.001$ ) was observed when Xinjiang samples were classified into four groups: WXJ (NChemur\_EMBA/NWafana\_EMBA/NSSG\_EMBA/W\_IA/W\_HE/SZGLK\_IA/SSPL\_HE/SHetian\_HE/SBZL\_HE), EXJ (N\_IA/SWJEZKL\_IA/E\_LBA), EIA (E\_IA/SRZG\_IA), and SEXiaohe\_BA (table S4). The EXJ group shared more affinity with NEA populations, WXJ groups clustered with western Steppe-related populations, and the EIA group shared more affinity with central Steppe and Europeans, as shown by the  $F_{ST}$  heatmap and DAPC plot (Fig. 2).

### Ancient Xinjiang comparisons with present-day populations

To determine the relationship that ancient Xinjiang populations share with present-day Eurasians, we compared Xinjiang individuals to four subgroups based on geographical locations: present-day NEA populations (pdNEA: Siberian and northern East Asian); SEA populations (pdSEA: southern East Asian and Southeast Asian); populations in and around Xinjiang (pdCA/XJ); and a combined European, Caucasus, and West Asian group ("pdEurWA") (figs. S5C and S6A). Among the BA populations, SEXiaohe\_BA shows the highest affinity for pdNEA-Siberians ( $F_{ST} = 0.12$ ,  $P = 0.00$ ) and E\_LBA for pdNEA-northern East Asian ( $F_{ST} < 0.02$ ,  $P > 0.05$  with Tu, Japanese, and Mongolian), while the northern and western Xinjiang EMBA shows the highest affinity for pdCA/XJ and pdEurWA ( $F_{ST} < 0.05$ ,  $P > 0.05$  with most populations), which are also observed in DAPC and haplogroup analyses (Fig. 2A and table S2). The IA and HE Xinjiang populations also show, in general, a high affinity for the populations in and around the Xinjiang region (pdCA/XJ) with very low  $F_{ST}$  values, demonstrating genetic connections between the IA, HE, and present-day populations (Fig. 4 and table S2). In addition, we see significant pdNEA connections with Xinjiang N\_IA ( $F_{ST} < 0.02$ ,  $P > 0.05$  with Japanese, Tu, Oroqen, and Han populations) and SWJEZKL\_IA ( $F_{ST} < 0.02$ ,  $P > 0.05$  in Tu and Mongolian) (fig. S6, B and C, and table S2).

The present-day Xinjiang populations from southwestern Xinjiang (29) include four populations: UyurXJ, KyrgyzXJ, SarTajikXJ, and WakTajikXJ. In DAPC, UyurXJ and KyrgyzXJ plot closer to the populations around Xinjiang (pdCA/XJ), and SarTajikXJ and WakTajikXJ cluster with pdEurWA populations (fig. S5, B and C). Our ancient Xinjiang samples clustered close to the UyurXJ and KyrgyzXJ, suggesting more East Asian affinities compared to the SarTajikXJ and WakTajikXJ, who show more affinity with Europeans and Iranians (fig. S5, B and C). In  $F_{ST}$  analysis, we also generally observe that Xinjiang EMBA and IA individuals show high genetic affinity with UyurXJ and KyrgyzXJ compared to the SarTajikXJ and WakTajikXJ, while Xinjiang HE samples show similar genetic

affinity for both of these groups (table S2). Thus, in summary, we find a genetic continuity from the ancient to present-day time periods whereby all the major ancestries from Siberia, Europe, East Asia, and West Central Asia are present in both ancient and present-day Xinjiang populations (28).

### DISCUSSION

Archaeological findings from Xinjiang have raised curiosity about its past population structure and the peopling of this region. Among competing hypotheses, most archaeologists support the view that ancient Xinjiang was a mixed confederation of people from both East and West Eurasia. The high mobility and admixture of people during the BA and IA periods around Xinjiang are supported by previous analyses with limited uniparental markers (18–20). However, the lack of ancient DNA from multiple regions and time periods has left Xinjiang's past population structure shrouded in mystery. Through our analyses, we have characterized the ancient population structure of Xinjiang and how it changed from the BA through the present day. These new data and results allowed us to propose a much more detailed and expansive admixture scenario.

The BA around Xinjiang was predominantly represented by western Steppe-related ancestries, which included the EMBA Yamnaya/Afanasievo cultures (28, 40, 42, 48). For instance, we find that the WSteppe\_EMBA populations cluster with individuals from the Songshugou site in northern Xinjiang (NSSG\_EMBA) in multiple haplogroups (U4, U5, H2, H6a, and W3). This is consistent with the Afanasievo-style relics at Songshugou (SSG) and the physical anthropology of an individual from this site (tomb M15) who shows European-like characteristics (49). We also find evidence for the influence of Chemurchek culture in BA Xinjiang, as suggested by the archaeological records of standing stone pillars with anthropomorphic figures around different cemeteries (50). The BA populations within Xinjiang were quite mixed genetically, as we found the presence of both East (NEA) and West (western Steppe-related) Eurasian mitochondrial haplogroups. Despite high admixture among the BA Xinjiang people, some unique genetic affinities are still observed. For example, western Steppe-related populations appear to have affected the northern and western Xinjiang populations (NWafana\_EMBA and NChemur\_EMBA) more notably than the eastern Xinjiang groups (E\_BA and E\_LBA), which showed more NEA connections. The NEA connection is consistent with the archaeologically hypothesized formation of the earliest known culture in BA eastern Xinjiang, the Tianshanbeilu culture (~3900 BP). It has been suggested that the Majiayao/Machang culture in Gansu Province, east of Xinjiang, formed the Tianshanbeilu culture in eastern Xinjiang and formed the Siba culture in the Hexi Corridor (51). Individuals from BA eastern Xinjiang (E\_BA; Tianshanbeilu culture) also showed physical similarities to populations from the Gan-Qing region (52). Consistent with these reports, we found that later LBA populations in eastern Xinjiang also had genetic connections to ancient Gan-Qing populations (GQQijia\_BA and GQKayue\_LBA). The presence of some West Eurasian-related haplogroups in eastern Xinjiang (both E\_BA and E\_LBA) is further consistent with the presence of burial forms, ornaments, and tools at the Tianshanbeilu site that share some West Eurasian features (53, 54), as well as some individuals that had European-like physical characteristics (55). SEXiaohe\_BA populations, however, showed more connections with ancient and present-day Siberian populations (Fig. 4A). The archaeological



findings from the Xiaohe site include wheat and millet grains, suggesting probable connections and exchanges from both West Asian along the IAMC and northern Chinese cultures (56, 57). This scenario likely suggests that migrations from both NEA and West Eurasia considerably affected the BA Xinjiang populations. All these western Steppe-related cultures, such as the Afanasievo, Chemurchek, and Botai, can represent a form of Altaic culture, which probably used the IAMC route to establish their presence in Xinjiang. In addition, the affinity of East and South Xinjiang BA populations with NEA and Siberians also suggests a greater influence from NEA and Siberia during the BA in Xinjiang.

The IA samples in Xinjiang continued to be more admixed with both East and West Eurasian lineages, but, similar to the BA, certain geographical regions showed different affinities for NEA and European populations. The IA populations from northern Xinjiang showed more ancient NEA connections, whereas the western and eastern populations (W\_IA and E\_IA) had an affinity for both NEA and European populations (Turan and western-related Steppe) and clustered with IA central Steppe populations (Fig. 4B). In IA western Xinjiang, the mixed affinity to East and West Eurasians may reflect the mixed cultural and physical similarities. The burial structure of the Early IA Suodunbulake culture in western Xinjiang is similar to the Central Asian Sapali and Wakeshi cultures in Amu Darya. The painted potteries of the Suodunbulake culture are more reminiscent of East Eurasian cultures, whereas individuals associated with the Suodunbulake have more European-like characteristics (58). The occupation of the Ili region by the Scythians in the early IA may also explain the shared affinity to both East and West Eurasian populations in IA western Xinjiang. The continued NEA connection observed in E\_IA reflects the continuity between the BA Tianshanbeilu and IA Yanbulake cultures in eastern Xinjiang, of which the latter was further affected by the Xindian culture from Gansu Province as well as the Chemurchek and Xintala cultures in Xinjiang (59).

Populations from southern Xinjiang (SZGLK\_IA) show more WSteppe\_MLBA and Turan connections, while the population from the high-elevation region of southwestern Xinjiang (SWJEZKL\_IA) shows more of an NEA connection. The regional preferences of some Xinjiang populations, particularly the differentiation between southwest and northeast Xinjiang, suggest that the IA was a very interactive time period. From the 200 BCE, the Silk Road passing through Xinjiang became influential in facilitating population migrations across Eurasia (60). We found that IA samples in northern Xinjiang (N\_IA) differed from the EMBA samples from this region in showing a closer affinity with NEA populations from the Hexi Corridor of the Gan-Qing region. Moreover, the IA samples from South Xinjiang (SZGLK\_IA) also had NEA lineages (C7b, D4i, and D4j1b), demonstrating a connection between northern China and the Tarim Basin, consistent with the influence of population movements along the Silk Road toward the Tarim Basin.

The HE populations continued to show both NEA and European lineages, reflecting the sustained high movement and admixture of people in this region. These HE populations are highly admixed, with societies of different cultural affinities living alongside one another (Fig. 4C). We also found that populations in both the N\_IA and W\_IA had Saka-related lineages, suggesting that the Saka people may have admixed with the IA populations in northern and western Xinjiang. The Xiongnu lineages (Fig. 3E and fig. S7D) found in the population from both eastern and western Tianshan (E\_IA and W\_IA) coincide with the westward expansion of the

Xiongnu population around the second or third century BCE (21). Hun mitochondrial lineages were observed in HE (but not IA) samples of southern Xinjiang [SBZL\_HE; 421 (95.4%) to 538 CE], coinciding with the invasion of the Hun into Scythian populations and the formation of the Hun-Scythians in the fourth to fifth century CE (21), suggesting a southward movement of this Hun tradition from the Ili region into southern Xinjiang in HE.

Xinjiang is associated with the extinct Indo-European Tocharian language, which was present from 500 to 900 CE in central Xinjiang based on ancient manuscripts (28, 61). In general, archaeologists view this language as being associated with Afanasievo-related people in Xinjiang (28). Our results for the BA sites suggest a complex scenario whereby the Xiaohe site in the Tarim Basin has a deep ancestral connection with ancient Siberian populations, whereas other Xinjiang EMBA populations from the north and west show a more Steppe EMBA (Afanasievo) connection. Thus, probably, the Tocharian language came into Xinjiang with populations associated with Steppe-related ancestry, such as the Afanasievo. However, this will require more sampling as Xiaohe (3900 to 3500 BP), which dates later than the other Xinjiang EMBA (~4500 BP), instead had a deep ANE connection with ancient Siberians. Khotanese, another ancient language, associated with the Indo-Iranian language family, was first observed in ancient documents at the Niya site (200 to 500 CE), Khotan, southern Tarim Basin (62), which is contemporaneous with our samples from this region (SSPL\_HE, 74 to 226 cal CE; SHetian\_HE, 138 to 345 cal CE; SBZL\_HE, 421 to 538 cal CE). The Khotanese language is associated with the expansion of the Sakas around 200 BCE into the Xinjiang region (63). We also observed the genetic affinity between many IA and HE Xinjiang populations with Sakas, suggesting its widespread presence in Xinjiang.

Therefore, the mitogenomic history of Xinjiang was heavily marked by western Steppe-related, central Steppe, northeastern Asian, and Turan introgression, and a confederation of different ancient populations is quite visible from the BA to HE periods. This admixture formed the foundation of present-day populations in Xinjiang, and future studies with ancient genomic data will reveal more admixture patterns in this region.

## MATERIALS AND METHODS

### Ancient sample collection

We collected samples from a total of 237 ancient human individuals from 41 sites (table S1). Their archaeological details are provided in the Supplementary Materials. Approval for their use was curated by co-authors and obtained with permission from the respective provincial archaeology institutes or universities that managed the samples. The permission and oversight were also provided by the institutional review board at the Institute of Vertebrate Paleontology and Paleoanthropology of the Chinese Academy of Sciences to study their ancient genomes.

### Ancient DNA (aDNA) extraction and library preparation

A total of 237 DNA extractions were obtained from the skeletal or dental remains of ancient Xinjiang individuals. Human remains were surface-cleaned and drilled for less than 100 mg of bone powder. Extraction was conducted in the aDNA clean room at the Institute of Vertebrate Paleontology and Paleoanthropology following strict aDNA standards (64). Single-stranded or double-stranded DNA was used to generate 237 libraries. We identified and retained

the cytosine-deaminated terminal sequences seen in typical aDNA and constructed a uracil-DNA glycosylase (“UDG”) partially treated, double-stranded (denoted as “DS\_half” in table S1) library (65). For library enrichment, we used an AccuPrime Pfx DNA enzyme for 35 cycles. P5 and P7 adapters were added to limit the contamination rate. We also used a NanoDrop 2000 spectrometer to monitor the DNA concentration.

### In-solution capture of mtDNA

The in-solution capture was accomplished by overlapping probes with DNA fragments and enriching the resulting libraries (66). The probes were synthesized on the basis of the human mitochondrial genome (67). It yielded an average mtDNA coverage of 460-fold (table S1).

### Sequencing and read alignment

The library pools were sequenced using an Illumina MiSeq instrument with paired  $2 \times 76$  base pair (bp) reads. We used *leeHom* (<https://github.com/grenaud/leeHom>) to trim adapters and merge sequences, with paired-end reads overlapping by at least 11 bp (68). Sequenced and merged reads with lengths of at least 30 bp were then mapped to the revised Cambridge Reference Sequence version 17 (rCRS) for mtDNA, using the *samse* command in the BWA v0.6.1 aligner (arguments used: `-n 0.01` and `-l16500`) (68, 69). 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. We used Haplogrep2 built on Phylotree Build 17 to call haplogroups for each sample (70, 71). We also sampled sequences randomly for each SNP covered at least once to recover allele information for each locus (table S1).

### Test for contamination

On the basis of the C to T transition rates estimated by MapDamage2.0 (72), we expect the true ancient DNA with proper transition rate. We also expect the mtDNA reads to match the consensus sequence better than 311 worldwide mtDNA genomes; we, therefore, used contamMix to detect present-day human mtDNA contamination (66). To avoid the damage pattern detected as contamination, we ignored the first and last five positions of the fragments. If the library had more than 4% of the fragments matched with other sequences that are better than the consensus, then we treated it as contaminated. However, for libraries Hetian\_M16A and Hetian\_M16B with substantial contamination but high coverage, a consensus was still retrieved by only using sequences with retained C to T substitutions on the first three positions at the 5′-end and the last three positions at the 3′-end of each fragment.

### Haplogroup analysis

The complete sequences of mtDNA were aligned with rCRS, based on PhyloTree v17 using MUSCLE (MUSCLE 3.3.31) software (70, 73). Haplogroups were called with haplogrep2 (70). We additionally used the DAPC package in R to conduct a haplogroup frequency-based DAPC and plotted PC1 and PC2 to illustrate haplogroup differences among populations (74). A heatmap based on the correlation of haplogroup frequencies was made using the heatmap function in R.

### Genetic distance analysis

We used the Arlequin software package to calculate genetic distances ( $F_{ST}$ ) between populations, which were visualized with the vegan

package’s metaMDS function in R (75). Heatmaps were also used to illustrate the statistical significance of clusters based on  $F_{ST}$ .

### Geographic distance and Mantel test

The geosphere package was used with R software to construct a geographic distance matrix from the geographic coordinates of each ancient Xinjiang group (37), excluding two extreme sites (SEXiaohe\_BA and NSSG\_EMBA). To estimate the correlation between the  $F_{ST}$  and geographic distance matrices, the *lm* function was used in R to fit the linear models (38). The regression results were plotted in R with its  $P$  and  $R^2$  values.

### AMOVA analysis

The Arlequin software package was also used to analyze the molecular genetic variance among groups, within each group, and among populations to assess population genetic structure (75).  $P$  values were calculated under 20,000 permutations to test whether the variance was statistically significant.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/14/eabd6690/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. S. Mu, Thirty years of archaeology in Xinjiang [in Chinese]. *Soc. Sci. Xinjiang* **3**, 86–97 (1985).
2. Y. Zhang, A brief introduction to Xinjiang archaeology [in Chinese]. *Archaeology* **6**, 3–13 (2002).
3. B. Wang, An analysis of the archaeological culture of the bronze age in Xinjiang. *Soc. Sci. Xinjiang* **4**, 50–61 (1985).
4. J. He, A preliminary study on the diet culture of the ancient inhabitants in Xinjiang from the prehistoric archaeology [in Chinese]. *Agric. Hist. China* **3**, 3–10 (2017).
5. Z. Yu, New archaeological discoveries in Xinjiang from 2006 to 2010 [in Chinese]. *Special Issue of Xinjiang* **4**, 44–51 (2011).
6. Institute of cultural relics and archaeology in Xinjiang [in Chinese]. Ten years of archaeology in Xinjiang. *Chinese Cultural Heritage* **1**, 98–111 (2007).
7. P. W. Jia, A. Betts, D. Cong, X. Jia, P. D. Dupuy, Adunqiaolu: New evidence for the Andronovo in Xinjiang, China. *Antiquity* **91**, 621–639 (2017).
8. D. Cong, W. Jia, B. Ailisen, X. Jia, D. Baola, Adunqiaolu: A new type of Bronze Age remains in the western Tianshan Mountains [in Chinese]. *The Western Regions Studies* **4**, 15–27 (2017).
9. X. Liu, P. J. Jones, G. M. Matuzeviciute, H. V. Hunt, D. L. Lister, T. An, N. Przelomska, C. J. Kneale, Z. Zhao, M. K. Jones, From ecological opportunism to multi-cropping: Mapping food globalisation in prehistory. *Quat. Sci. Rev.* **206**, 21–28 (2019).
10. P. N. Doumani Dupuy, P. W. Jia, A. Betts, D. Cong, Pots and potters of the Bronze Age of north-west Xinjiang. *Antiquity* **371**, 1231–1248 (2019).
11. A. Betts, P. Jia, I. Abuduresule, A new hypothesis for early Bronze Age cultural diversity in Xinjiang, China. *Archaeol. Res. Asia* **17**, 204–213 (2019).
12. Y. Li, Agriculture and palaeoeconomy in prehistoric Xinjiang, China (3000–200 BCE). *Veg. Hist. Archaeobot.* **10**, (2020).
13. T. Shui, The theoretical problems on the development of prehistoric civilization in the western regions [in Chinese]. *The Western Regions Studies* **4**, 44–51 (2010).
14. T. Long, C. Leipe, G. Jin, M. Wagner, R. Guo, O. Schroder, P. E. Tarasov, The early history of wheat in China from 14C dating and Bayesian chronological modelling. *Nat. Plants* **4**, 272–279 (2018).
15. Y. Fang, A preliminary study on the prehistoric bronze mirror in Xinjiang [in Chinese]. *Chang’an Academic Journal: Philosophy and Social Sciences* **1**, 14–17 (2018).
16. C. Zhang, M. Liu, Views on the study of prehistoric bronzes in Xinjiang [in Chinese]. *Journal of Urumqi Vocational University* **1**, 40–44 (2011).
17. T. Shui, The formation and influence of the interaction circle in the Altai mountains during the Bronze Age – The discovery of China as the main line [in Chinese]. *The Ancient Civilization* **10**, 262–267 (2016).
18. S. Gao, Y. Zhang, D. Wei, H. Li, Y. Zhao, Y. Cui, H. Zhou, Ancient DNA reveals a migration of the ancient Di-Qiang populations into Xinjiang as early as the early Bronze Age. *Am. J. Phys. Anthropol.* **157**, 71–80 (2015).

19. C. Li, C. Ning, E. Hagelberg, H. Li, Y. Zhao, W. Li, I. Abuduresule, H. Zhu, H. Zhou, Analysis of ancient human mitochondrial DNA from the Xiaohu cemetery: Insights into prehistoric population movements in the Tarim Basin, China. *BMC Genet.* **16**, 78–88 (2015).
20. C. Li, H. Li, Y. Cui, C. Xie, D. Cai, W. Li, V. H. Mair, Z. Xu, Q. Zhang, I. Abuduresule, L. Jin, H. Zhu, H. Zhou, Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol.* **8**, 15–26 (2010).
21. P. de Barros Damgaard, N. Marchi, S. Rasmussen, M. Peyrot, G. Renaud, T. Korneliusson, J. V. Moreno-Mayar, M. W. Pedersen, A. Goldberg, E. Usmanova, N. Baimukhanov, V. Loman, L. Hedeager, A. G. Pedersen, K. Nielsen, G. Afanasiev, K. Akmatov, A. Aldashev, A. Alpaslan, G. Baimbetov, V. I. Bazaliiskii, A. Beisenov, B. Boldbaatar, B. Boldgiv, C. Dorzhu, S. Ellingvag, D. Erdenebaatar, R. Djani, E. Dmitriev, V. Evdokimov, K. M. Frei, A. Gromov, A. Goryachev, H. Hakonarson, T. Hegay, Z. Khachatryan, R. Khakhanov, E. Kitov, A. Kolbina, T. Kubatbek, A. Kukushkin, I. Kukushkin, N. Lau, A. Margaryan, I. Merkyte, I. V. Mertz, V. K. Mertz, E. Mijdorj, V. Moiseyev, G. Mukhtarova, B. Nurmukhanbetov, Z. Orozbekova, I. Panyushkina, K. Pieta, V. Smrčka, I. Shevnina, A. Logvin, K. G. Sjögren, T. Štolcová, K. Tashbaeva, A. Tkachev, T. Tulegenov, D. Voyakin, L. Yepiskoposyan, S. Undrakhbold, V. Varfolomeev, A. Weber, N. Krادين, M. E. Allentoft, L. Orlando, R. Nielsen, M. Sikora, E. Heyer, K. Kristiansen, E. Willerslev, 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
22. J. Yu, J. Ma, A preliminary study on the Dongtaleda cemetery of Habahe County, Xinjiang [in Chinese]. *Cultural Relics* **3**, 53–57 (2013).
23. L. A. Sokolova, Okunev cultural tradition in the stratigraphic aspect. *Archaeol. Ethnol. Anthropol. Eurasia* **30**, 41–51 (2007).
24. H. Yu, M. A. Spyrou, M. Karapetian, S. Shnaider, R. Radzevičiūtė, K. Nägele, G. U. Neumann, S. Penseke, J. Zech, M. Lucas, P. LeRoux, P. Roberts, G. Pavlenok, A. Buzhilova, C. Posth, C. Jeong, J. Krause, Paleolithic to Bronze Age Siberians reveal connections with first Americans and across Eurasia. *Cell* **6**, 1232–1245.e20 (2020).
25. C. Hollard, V. Zvenigorosky, A. Kovalev, Y. Kiryushin, A. Tishkin, I. Lazaretov, E. Crubézy, B. Ludes, C. Keyser, New genetic evidence of affinities and discontinuities between bronze age Siberian populations. *Am. J. Phys. Anthropol.* **167**, 97–107 (2018).
26. C. Ning, C.-C. Wang, S. Gao, Y. Yang, X. Zhang, X. Wu, F. Zhang, Z. Nie, Y. Tang, M. Robbeets, J. Ma, J. Krause, Y. Cui, Ancient genomes reveal Yamnaya-related Ancestry and a potential source of Indo-European Speakers in Iron Age Tianshan. *Curr. Biol.* **29**, 2526–2532.e4 (2019).
27. P. Miao, The theory of Xiongnu ruling the western region [in Chinese]. *The Western Regions Studies* **2**, 1–9 (2016).
28. V. M. Narasimhan, N. Patterson, P. Moorjani, N. Rohland, R. Bernardos, S. Mallick, I. Lazaridis, N. Nakatsuka, I. Olalde, M. Lipson, A. M. Kim, L. M. Olivieri, A. Coppa, M. Vidale, J. Mallory, V. Moiseyev, E. Kitov, J. Monge, N. Adamski, N. Alex, N. Broomandkoshbacht, F. Candilio, K. Callan, O. Cheronet, B. J. Culleton, M. Ferry, D. Fernandes, S. Freilich, B. Gamarra, D. Gaudio, M. Hajdinjak, E. Harney, T. K. Harper, D. Keating, A. M. Lawson, M. Mah, K. Mandl, M. Michel, M. Novak, J. Oppenheimer, N. Rai, K. Sirak, V. Slon, K. Stewardson, F. Zalzal, Z. Zhang, G. Akhatov, A. N. Bagashev, A. Bagnera, B. Baitanayev, J. Bendezu-Sarmiento, A. A. Bissebaev, G. L. Bonora, T. T. Charyginov, T. Chikisheva, P. K. Dashkovskiy, A. Derevianko, M. Dobeš, K. Douka, N. Dubova, M. N. Duisengali, D. Enshin, A. Epimakhov, A. V. Fribus, D. Fuller, A. Goryachev, A. Gromov, S. P. Grushin, B. Hanks, M. Judd, E. Kazizov, A. Khokhlov, A. P. Krygin, E. Kupriyanova, P. Kuznetsov, D. Luiselli, F. Maksudov, A. M. Mamedov, T. B. Mamirov, C. Meiklejohn, D. C. Merrett, R. Micheli, O. Mochalov, S. Mustafokulov, A. Nayak, D. Pettener, R. Potts, D. Razhev, M. Rykun, S. Sarno, T. M. Savenkova, K. Sikhymbaeva, S. M. Slepchenko, O. A. Soltobayev, N. Stepanova, S. Svyatko, K. Tabaldiev, M. Teschler-Nicola, A. A. Tishkin, V. V. Tkachev, S. Vasilyev, P. Velemínský, D. Voyakin, A. Yermolayeva, M. Zahir, V. S. Zubkov, A. Zubova, V. S. Shinde, C. Lalueza-Fox, M. Meyer, D. Anthony, N. Boivin, K. Thangaraj, D. J. Kennett, M. Frachetti, R. Pinhasi, D. Reich, The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
29. M. Peng, W. Xu, J. Song, X. Chen, X. Sulaiman, L. Cai, H. Liu, S. Wu, Y. Gao, N. T. Abdulloevich, M. E. Afanasevna, K. B. Khudoidodov, X. Chen, W. Yang, M. Wu, G. Li, X. Yang, A. Rakha, Y. Yao, H. Upur, Y. Zhang, Mitochondrial genomes uncover the maternal history of the Pamir populations. *Eur. J. Hum. Genet.* **26**, 124–136 (2018).
30. Q. Feng, Y. Lu, X. Ni, K. Yuan, Y. Yang, X. Yang, C. Liu, H. Lou, Z. Ning, Y. Wang, D. Lu, C. Zhang, Y. Zhou, M. Shi, L. Tian, X. Wang, X. Zhang, J. Li, A. Khan, Y. Guan, K. Tang, S. Wang, S. Xu, Genetic history of Xinjiang's Uyghurs suggests Bronze Age multiple-way contacts in Eurasia. *Mol. Biol. Evol.* **34**, 2572–2582 (2017).
31. X. Wu, Archaeological excavation of Jierzankale Cemetery in Taxkorgan of Xinjiang in 2013 [in Chinese]. *The Western Regions Studies* **1**, 124–127 (2014).
32. M. Wang, X. Zhang, X. Wu, Preliminary study on human bones in Jierzankale Cemetery, Taxkorgan, Xinjiang [in Chinese]. *Northern Cultural Relics* **4**, 42–52 (2019).
33. H. Shao, An introduction to Zaghunluq Culture in Xinjiang [in Chinese]. *Research of China's Frontier Archaeology* **7**, 170–183 (2008).
34. C. Xie, C. Li, Y. Cui, D. Cai, Mitochondrial DNA analysis of ancient Sampula population in Xinjiang [in Chinese]. *Prog. Nat. Sci.* **17**, 927–933 (2007).
35. X. Xiao, B. Zheng, Excavation of Sampula tomb in Luopu County, Xinjiang [in Chinese]. *The Western Regions Studies* **1**, 42–46 (2000).
36. X. Hu, A. Niyazi, A new archaeological research in Bizili Cemetery, Luopu County, Xinjiang [in Chinese]. *The Western Regions Studies* **1**, 144–146 (2017).
37. C. F. F. Karney, Algorithms for geodesics. *J. Geodesy.* **87**, 43–55 (2013).
38. G. N. Wilkinson, C. E. Rogers, Symbolic descriptions of factorial models for analysis of variance. *J. R. Stat. Soc.* **22**, 392–399 (1973).
39. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. Rojo Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
40. P. B. Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamovic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaitsev, V. Merz, A. Biddanda, I. Merz, V. Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediyah, I. Ullah, K.-G. Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A. Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. M. Kenzie, A. Margaryan, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E. Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
41. M. Derenko, B. Malyarchuk, T. Grzybowski, G. Denisova, U. Rogalla, Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in Northern Asia. *PLOS ONE* **5**, e15214 (2010).
42. C. Jeong, S. Wilkin, T. Amgalantugs, A. S. Bouwman, W. T. T. Taylor, R. W. Hagan, S. Bromage, S. Tsolmon, C. Trachsel, J. Grossmann, J. Littleton, C. A. Makarewicz, J. Krigbaum, M. Burri, A. Scott, G. Davaasambuu, J. Wright, F. Irmer, E. Myagmar, N. Boivin, M. Robbeets, F. J. Rühli, J. Krause, B. Frohlich, J. Hendy, C. Warinner, Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11248–E11255 (2018).
43. M. Ding, T. Wang, A. M.-S. Ko, H. Chen, H. Wang, G. Dong, H. Lu, W. He, S. Wangdue, H. Yuan, Y. He, L. Cai, Z. Chen, G. Hou, D. Zhang, Z. Zhang, P. Cao, Q. Dai, X. Feng, M. Zhang, H. Wang, M. A. Yang, Q. Fu, Ancient mitogenomes show plateau populations from last 5200 years partially contributed to present-day Tibetans. *P. R. Soc. B Biol. Sci.* **287**, 20192968 (2020).
44. C. Mittnik, C.-C. Wang, S. Pfrengle, M. Daubaras, G. Zariņa, F. Hallgren, R. Allmāe, V. Khartanovich, V. Moiseyev, M. Törv, A. Furtwängler, A. A. Valtueña, M. Feldman, C. Economou, M. Oinonen, A. Vasks, E. Balanovska, D. Reich, R. Jankauskas, W. Haak, S. Schiffels, J. Krause, The genetic prehistory of the Baltic Sea region. *Nat. Comm.* **9**, 442 (2018).
45. H. Sahakyan, B. H. Kashani, R. Tamang, A. Kushniarevich, A. Francis, M. D. Costa, A. K. Pathak, Z. Khachatryan, I. Sharma, M. van Oven, J. Parik, H. Hovhannisyán, E. Metspalu, E. Pennarun, M. Karmin, E. Tamm, K. Tambets, A. Bahmanimehr, T. Reisberg, M. Reidla, A. Achilli, A. Olivieri, F. Gandini, U. A. Perego, N. Al-Zahery, M. Houshmand, M. H. Sanati, P. Soares, E. Rai, J. Šarac, T. Šarič, V. Sharma, L. Pereira, V. Fernandes, V. Černý, S. Farjadian, D. P. Singh, H. Azakli, D. Üstek, N. Ekomasova, I. Kutuev, S. Litvinov, M. Bermisheva, E. K. Khusnutdinova, N. Rai, M. Singh, V. K. Singh, A. G. Reddy, H.-V. Tolk, S. Cvjetan, L. B. Lauc, P. Rudan, E. N. Michalodimitrakis, N. P. Anagnou, K. I. Pappa, M. V. Golubenko, V. Orekhov, S. A. Borinskaya, K. Kaldma, M. A. Schauer, M. Simionescu, V. Gusar, E. Grechanina, P. Govindaraj, M. Voevoda, L. Damba, S. Sharma, L. Singh, O. Semino, D. M. Behar, L. Yepiskoposyan, M. B. Richards, M. Metspalu, T. Kivisild, K. Thangaraj, P. Endicott, G. Chaubey, A. Torroni, R. Villemes, Origin and spread of human mitochondrial DNA haplogroup U7. *Sci. Rep.* **7**, 46044 (2017).
46. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
47. M. D. Frachetti, Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr. Anthropol.* **53**, 2–38 (2012).
48. C. Jeong, K. Wang, S. Wilkin, W. T. T. Taylor, B. K. Miller, J. H. Bemmman, R. Stahl, C. Chiorelli, F. Knolle, S. Ulziibayar, D. Khatanbaatar, D. Erdenebaatar, U. Erdenebat, A. Ochir, G. Ankhsanaa, C. Vanchigdash, B. Ochir, C. Munkhbayar, D. Tumen, A. Kovalev, N. Krادين, B. A. Bazarov, D. A. Miyagashev, P. B. Konovalov, E. Zhambaltarova, A. V. Miller, W. Haak, S. Schiffels, J. Krause, N. Boivin, M. Erdene, J. Hendy, C. Warinner, A dynamic 6,000-year genetic history of Eurasia's Eastern Steppe. *Cell* **183**, 890–904.e29 (2020).

49. Institute of cultural relics and archaeology in Xinjiang, Archaeological excavation of Songshuguo site, Hoboksar County, Ili region, Xinjiang [in Chinese]. *Xinjiang Cultural Relics* **1**, (2018).
50. Institute of cultural relics and archaeology in Xinjiang, The excavation briefing of Tuoganbai NO.2 Cemetery in Habahe, Xinjiang [in Chinese]. *Cultural Relics* **12**, 18–28 (2014).
51. J. Han, Study on painted pottery before the Silk Road [in Chinese]. *Journal of Archaeology and museology* **1**, 20–32 (2018).
52. D. Wei, *Study on the Migrant and Communication of Ancient People from Bronze Age to Early Iron Age in Kumul Region, Xinjiang* [in Chinese] (Science Press, 2017).
53. H. Shao, *Cultural Evolution of Prehistoric Xinjiang and the Relationship with Cultures Around Xinjiang* [in Chinese] (Science Press, 2018).
54. J. Han, The stage and pedigree of Bronze Age and Iron Age cultures in Xinjiang [in Chinese]. *Xinjiang Cultural Relics* **3**, (2005).
55. B. Wang, The racial study on unearthed human cranial from Tianshanbeilu site, Kumul region, Xinjiang [in Chinese]. *Journal of Xinjiang Normal University* **1**, 97–107 (2013).
56. C. Li, D. L. Lister, H. Li, Y. Xu, Y. Cui, M. A. Bower, M. K. Jones, H. Zhou, Ancient DNA analysis of desiccated wheat grains excavated from a bronze age cemetery in Xinjiang. *J. Archaeol. Sci.* **38**, 115–119 (2011).
57. J. Yu, Pre-historical millet crops from Xinjiang [in Chinese]. *The Western Regions Studies* **3**, 71–75 (2012).
58. L. Chen, Study on unearthed human cranial from Suodunbulake site, Qapqal County, Ili region, Xinjiang [in Chinese]. *Archaeology* **7**, 79–94 (2003).
59. J. Han, *The Bronze Age and Iron Age cultures in Xinjiang* [in Chinese] (Cultural Relics Press, 2007).
60. V. Hansen, *The Silk Road: A New History* (Oxford Univ. Press, 2012).
61. W. Guo, *The Problem of Tocharian in the View of Archaeological Field* [in Chinese] (Science Press, 2015), pp. 373–403.
62. Y. Yang, The cultural exchange of central Plain reflected by Niya site [in Chinese]. *The Western Regions Studies* **2**, (1999).
63. H. W. Bailey, Saka Studies: The ancient Kingdom of Khotan. *Iran* **8**, 65–72 (1970).
64. M. T. P. Gilbert, H. J. Bandelt, M. Hofreiter, I. Barnes, Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005).
65. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic. Acids. Res.* **40**, e3 (2012).
66. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013).
67. R. M. Andrews, I. Kubacka, P. F. Chinnery, R. N. Lightowler, D. M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
68. R. Gabriel, S. Udo, K. Janet, leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic. Acids. Res.* **42**, e141 (2014).
69. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic. Acids. Res.* **44**, W58–W63 (2016).
71. M. V. Oven, PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.* **5**, e392–e394 (2015).
72. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
73. R. C. Edgar, MUSCLE: Multiple sequence alignment with improved accuracy and speed. *Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004* (Stanford, CA, USA, 2004), pp. 1792–1797.
74. B. François, D. Sébastien, J. Thibaut, Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
75. L. Excoffier, H. E. L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
76. National Genomics Data Center Members and Partners, Database resources of the national genomics data center in 2020. *Nucleic. Acids. Res.* **48**, D24–D33 (2020).
77. Institute of cultural relics and archaeology in Xinjiang, Archaeological investigation and excavation of Xiaohe cemetery, southern Xinjiang in 2002 [in Chinese]. *Research of China's Frontier Archaeology* **3**, 338–398 (2004).
78. Institute of cultural relics and archaeology in Xinjiang, Brief excavation of Xiaohe cemetery in 2003, Lop Nur, Xinjiang [in Chinese]. *Culture Relics* **10**, 4–42 (2007).
79. M. Lin, The origin and migrant of Tocharians [in Chinese]. *The Western Regions Studies* **3**, 9–23 (2002).
80. Y. Nie, H. Zhu, W. Li, Y. Abuduresule, Anthropological features of the skulls of ancient populations in Xiaohe cemetery, Lop Nur, Xinjiang [in Chinese]. *The western regions Studies* **3**, 115–125 (2010).
81. Institute of Archaeology, Chinese Academy of Social Science, Report on the archaeological excavation of Jierzankale Cemetery in Taxkorgan of Xinjiang [in Chinese]. *Acta Archaeologica Sinica* **2**, 229–252 (2012).
82. Xinjiang Museum, Report on excavation of Zaghunluq NO.1, Qiemo County, Xinjiang [in Chinese]. *Acta Archaeologica Sinica* **1**, 89–136 (2003).
83. Institute of cultural relics and archaeology in Xinjiang, Report on archaeological excavation of the NO.1 Cemetery group of Ayituoahan in Habahe County [in Chinese]. *Xinjiang Cultural Relics* **2**, 19–39 (2017).
84. S. Li, The Afanasievo Culture from the discovery of the Ayituoahan NO.1 Cemetery in Xinjiang [in Chinese]. *Cultural Relics in Xinjiang* **1–2**, 105–121 (2018).
85. A. Chen, The characteristics and cultural origin of Chemurchek stone people [in Chinese]. *Research on Turpan* **2**, 105–112 (2012).
86. J. Yu, Archaeological records of Bolati tombs [in Chinese]. *Popular Archaeology* **6**, 73–77 (2013).
87. W. Guo, New archaeological harvest in Chagangole township in 2015, Qinghe County, Xinjiang [in Chinese]. *The Western Regions Studies* **1**, 135–137 (2016).
88. Institute of Cultural Relics and Archaeology in Xinjiang, The excavation briefing on Dongtaleda Cemetery in Habahe County, Xinjiang [in Chinese]. *Cultural Relics* **3**, 4–14 (2013).
89. Institute of Archaeology, Xinjiang Academy of Social Sciences, The excavation briefing on Kermuqi cemeteries [in Chinese]. *Cultural Relics* **1**, 23–32 (1981).
90. H. Shao, thesis, Jilin University (2007).
91. C. B. Jixieliefu, *The Ancient History of Southern Siberia* [in Chinese] (Xinjiang People's Press, 1981).
92. G. Tian, S. Guo, Xiongnu relics found in Aluchaideng, Inner Mongolia [in Chinese]. *Archaeology* **4**, 333–338 (1980).
93. S. Ye, The Jade Road in Steppe and the spread of Onyx beads in China, a from 2000 to 1000 BCE [in Chinese]. *Inner Mongolia Social Sciences* **4**, 137–147 (2018).
94. J. Yu, X. Hu, Y. Yu, Y. Liu, X. He, J. Zhang, N. Alipujiang, W. Tong, A brief report on the excavation of the tomb in Tuwaxincun, Kanas River, Burqi County, Xinjiang [in Chinese]. *Cultural Relics* **7**, 4–16 (2014).
95. B. Wang, X. Qi, The type analysis of stone people in Xinjiang [in Chinese]. *The western regions Studies* **4**, 67–76 (1995).
96. X. Chang, Brief report on Nanwan cemetery, Barkol County, Xinjiang [in Chinese]. *Xinjiang Cultural Relics* **1**, (1985).
97. X. He, Tomb M95 of Nanwan site, Barkol County, Xinjiang [in Chinese]. *Archaeology and Cultural Relics* **5**, 6–8 (1987).
98. School of Cultural Heritage, Northwest University, Brief report on excavation of Haiziyan site, Balkol County, Xinjiang [in Chinese]. *Cultural Relics* **12**, (2020).
99. J. Wang, F. Zhang, M. Ren, Report on the Dongheigou site in Barkol County, Xinjiang, during 2006–2007. *Archaeology* **1**, 3–27 (2009).
100. Z. Xin, A. T. Rodrigues, Y. Yue, W. Jianxin, M. Jian, R. Meng, Y. Jing, Y. Dongya, Ancient DNA analysis of ancient horse remains from Shirenzigou site, Xinjiang. *Quaternary Sci.* **34**, 187–195 (2014).
101. T. Shui, *A Comparative Study on Various Cultures in the Bronze Age of Xinjiang—The Historical Process of the Early Cultural Exchange between East and the West Eurasian* [in Chinese] (Science Press, 2002).
102. Q. Zhang, X. Chang, G. Liu, Stable isotope analysis of human bones unearthed from Tianshanbeilu Cemetery in Kumul City, Xinjiang. *The Western Regions Studies* **2**, 38–43 (2010).
103. W. Dong, thesis, Lanzhou University (2016).
104. Y. Wang, Z. Dang, The new archaeological discovery of Aisikexiaernan site in Kumul region, Xinjiang [in Chinese]. *The western regions Studies* **2**, 134–137 (2011).
105. Z. Tang, A typological comparative study of materials from the Heigouliang and Dongheigou Cemeteries in Barkol County, Xinjiang [in Chinese]. *Archaeology and Heritage* **5**, 51–56 (2010).
106. Q. Zhang, X. Chang, G. Liu, The dietary structure of the human bones unearthed from Heigouliang site, Barkol County, Xinjiang [in Chinese]. *The Western Regions Studies* **3**, 45–49 (2009).
107. H. Liu, B. Teer, X. Wang, B. Guan, H. Liu, Archaeological harvest and preliminary understanding of Dunna Expressway–Nilka section in Ili, Xinjiang Province [in Chinese]. *The Western Regions Studies* **3**, 137–139 (2018).
108. Institute of cultural relics and archaeology in Xinjiang, Excavation of ancient tombs along Dunna Expressway–Nilka section in Ili, Xinjiang [in Chinese]. *Archeology* **12**, (2020).
109. Q. Ruan, W. Hu, J. Zhang, The report on the excavation of the Wutulan Cemetery in Nilka, Xinjiang [in Chinese]. *Cultural Relics* **12**, 50–63 (2014).
110. Xinjiang Institute of Culture Relics and Archaeology, Excavation of the Wutulan Cemetery, Nilka, Xinjiang [in Chinese]. *Cultural Relics* **2**, 4–17 (2015).

111. Q. Ruan, A. Niyazi, Y. Wang, A brief report on the excavation of Tangbalesayi Cemetery in Nilka County, Ili, Xinjiang [in Chinese]. *Cultural Relics* **5**, 13–22 (2012).
112. Institute of cultural relics and archaeology in Xinjiang, Excavation of Kuokesuxi NO.2 Cemetery in Tekes County, Xinjiang [in Chinese]. *Archaeology* **9**, 3–16 (2012).
113. Y. Wang, Q. Ruan, The Jirentaigoukou site, Nilka County, Xinjiang [in Chinese]. *Public archaeology* **7**, 57–70 (2017).
114. Y. Wang, A new harvest from the archaeological work in Jirentaigoukou, Nilka County, Xinjiang in 2015 [in Chinese]. *The western regions Studies* **1**, 132–134 (2016).
115. Y. Wang, X. Yuan, Q. Ruan, The archaeological harvest and preliminary understanding of Jirentaigoukou site in Nilka County, Xinjiang from 2015 to 2018 [in Chinese]. *The western regions Studies* **1**, 133–138 (2019).
116. Institute of cultural relics and archaeology in Xinjiang, Report on the excavation of Ayousaigoukou site, and Kalaaoiy site, Nilka County, Ili, Xinjiang [in Chinese]. *Xinjiang Cultural Relics* **2**, (2012).
117. Institute of cultural relics and archaeology in Xinjiang, A brief report on the excavation of Ayousaigoukou site, Nilka County, Ili, Xinjiang [in Chinese]. *Cultural Relics* **5**, 13–22 (2012).

**Acknowledgments:** We would like to thank M. A. Yang, E. A. Bennett, and Y. Liu for comments as well as archaeological teams from Xinjiang and Xi'an. We would also like to thank the reviewers, which included V. M. Narasimhan and two anonymous reviewers, who provided valuable suggestions that helped to further improve the manuscript. **Funding:** This work was supported by National Key R&D Program of China (2016YFE0203700), the Chinese Academy of Sciences (CAS, XDB26000000), National Natural Science Foundation of China (91731303, 41925009, 41672021, and 41630102), CAS (XDA1905010 and QYZDB-SSW-DQC003), Tencent

Foundation through the XPLOER PRIZE, and the Howard Hughes Medical Institute (grant no. 55008731). **Author contributions:** Q.F. designed the research project. Q.F., V.K., W.L., and J.M. managed the project. X.H., Q.R., W.G., Xinhua Wu, J.Y., B.W., Z.T., N.A., J.Z., X.C., Y.T., M.R., Xiaohong Wu, M.Z., W.H., and T.W. assembled archaeological materials and dating. Q.F., P.C., R.Y., F.L., X.F., Q.D., and W.P. performed or supervised wet laboratory work. Q.F. and X.F. did the data processing and quality control. W.W. and M.D. analyzed the data. W.W., M.D., J.D.G., Y.W., B.M., V.K., and Q.F. wrote the manuscript. All authors discussed, critically revised, and approved the final version of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The mitochondrial genome consensus reported in this paper has been deposited in the Genome Warehouse in the National Genomics Data Center (76), Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences, under BioProject accession number PRJCA002948 that is publicly accessible at <https://bigd.big.ac.cn/gwh>. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 6 July 2020

Accepted 11 February 2021

Published 31 March 2021

10.1126/sciadv.abd6690

**Citation:** W. Wang, M. Ding, J. D. Gardner, Y. Wang, B. Miao, W. Guo, X. Wu, Q. Ruan, J. Yu, X. Hu, B. Wang, X. Wu, Z. Tang, A. Niyazi, J. Zhang, X. Chang, Y. Tang, M. Ren, P. Cao, F. Liu, Q. Dai, X. Feng, R. Yang, M. Zhang, T. Wang, W. Ping, W. Hou, W. Li, J. Ma, V. Kumar, Q. Fu, Ancient Xinjiang mitogenomes reveal intense admixture with high genetic diversity. *Sci. Adv.* **7**, eabd6690 (2021).