# Ancient Mitochondrial Genomes Reveal Extensive Genetic Influence of the Steppe Pastoralists in Western Xinjiang

Chao Ning[1,2†], Hong-Xiang Zheng[3†], Fan Zhang[1†], Sihao Wu[1], Chunxiang Li[1], Yongbin Zhao[4], Yang Xu[1], Dong Wei[5], Yong Wu[6], Shizhu Gao[7]*, Li Jin[3]* and Yinqiu Cui[1]*

[1] School of Life Sciences, Jilin University, Changchun, China, [2] Max Planck Institute for the Science of Human History, Jena, Germany, [3] State Key Laboratory of Genetic Engineering, School of Life Sciences, and Human Phenome Institute, Fudan University, Shanghai, China, [4] College of Life Science, Jilin Normal University, Siping, China, [5] School of Archaeology, Jilin University, Changchun, China, [6] Xinjiang Cultural Relics and Archaeology Institute, Urumchi, China, [7] School of Pharmaceutical Sciences, Jilin University, Changchun, China

The population prehistory of Xinjiang has been a hot topic among geneticists, linguists, and archaeologists. Current ancient DNA studies in Xinjiang exclusively suggest an admixture model for the populations in Xinjiang since the early Bronze Age. However, almost all of these studies focused on the northern and eastern parts of Xinjiang; the prehistoric demographic processes that occurred in western Xinjiang have been seldomly reported. By analyzing complete mitochondrial sequences from the Xiabandi (XBD) cemetery (3,500–3,300 BP), the up-to-date earliest cemetery excavated in western Xinjiang, we show that all the XBD mitochondrial sequences fall within two different West Eurasian mitochondrial DNA (mtDNA) pools, indicating that the migrants into western Xinjiang from west Eurasians were a consequence of the early expansion of the middle and late Bronze Age steppe pastoralists (Steppe_MLBA), admixed with the indigenous populations from Central Asia. Our study provides genetic links for an early existence of the Indo-Iranian language in southwestern Xinjiang and suggests that the existence of Andronovo culture in western Xinjiang involved not only the dispersal of ideas but also population movement.

Keywords: mitochondrial genome, ancient DNA, Eurasian Steppe, Silk Road, Andronovo

## INTRODUCTION

Recent archaeogenetic studies showed that the expansion of western steppe herders (WSHs) had a marked impact on the demographic, cultural, social and linguistic development since the third millennium BCE on the Eurasian continent (Allentoft et al., 2015; Haak et al., 2015; Damgaard et al., 2018a; Jeong et al., 2018, 2020; Narasimhan et al., 2019; Wang C. C. et al., 2021). One of the earliest representatives, known as the Yamnaya culture (ca. 3,300–2,700 BCE) from the Pontic–Caspian steppe migrated into Europe and Asia, bringing with them metallurgy, animal herding skills, and possibly the Indo-European languages (Frachetti, 2009; Allentoft et al., 2015; Haak et al., 2015). By the middle and late Bronze Age, the Sintashta culture (ca. 2,200–1,800 BCE) arose near the Urals and succeeded a majority of ancestry from the preceding Yamnaya culture. It carried a similar

genetic profile with the Srubnaya and the Andronovo cultures that spread over a large part of the Eurasia landmass, extending westward into Europe, southward into Central Asia and the India subcontinent, and eastward into the Mongolian Plateau (Allentoft et al., 2015; Haak et al., 2015; Damgaard et al., 2018b; Narasimhan et al., 2019; Jeong et al., 2020; Wang C. C. et al., 2021). A number of studies provided the evidence that the steppe cultures from western Eurasia had also integrated into the early Bronze Age cultures of western China. A recent archaeobotanical study showed that both wheat and barley had already spread to the Altai Mountains as early as 5,200 years ago (Zhou et al., 2020). Additionally, domesticated sheep and cattle were also observed in the prehistoric cultures of northwestern China (e.g., Majiayao culture, 3,550–2,850 BC; Qijia, 2,450–1,650 BC) (Fu et al., 2009). The cultural influences from WSHs suggested that ancient mobile pastoralists had played an extremely significant role in the prehistoric *trans*-Eurasian exchanges and the formation of agropastoralism.

Located at the intersection of the ancient "Silk Road," Xinjiang has played an important role in bridging the exchanges of cultures, goods, languages, and population movements (Wood, 2002). A recent genome-wide study on 951 Uyghurs in Xinjiang revealed a complex demographic history of the present-day populations in this region. Four major ancestral components were identified, namely, European, South Asian, Siberian, and East Asian (Feng et al., 2017). Two waves of admixtures were further characterized, with the first wave dating back as early as 3,750 years ago (Feng et al., 2017). However, human populations always underwent frequent population migrations, admixtures, and replacements, which made it difficult to reflect the true ancestral components and population dynamics using extant population data alone. The high level of genetic diversity of present-day Xinjiang people was likely a result of recent admixture events. The opening of the "Silk Road" made the exchanges of different populations in Xinjiang more frequent than ever. In contrast, ancient DNA study has been proven to be a powerful tool to reconstruct human prehistory by providing direct tests on samples from a certain period. Previous genetic studies have delineated that modern and ancient Xinjiang populations had maternal genetic affinities with both the eastern and western Eurasians, displaying high genetic diversity and admixture (Yao et al., 2004; Li et al., 2010; Zhang et al., 2010; Zheng et al., 2017; Wang W. et al., 2021). A recent paleogenomic study on the Iron Age Shirenzigou individuals from the eastern Tianshan mountains further confirmed the previous observations and characterized that the West Eurasian ancestry was likely to be related to the Early Bronze Age steppe pastoralists such as Yamnaya and/or Afanasievo than the chronologically more recent Sintashta and Andronovo cultures (Ning et al., 2019). Wang W. et al. (2021) retrieved the whole mitochondrial genomes of ancient Xinjiang populations from the Bronze Age to Historic Era. Their results revealed that the Bronze Age Xinjiang populations had genetic affinities with Steppe-related and Northeastern Asian populations (Wang W. et al., 2021). All of the above studies had proven the very complex demographic landscape of the ancient Xinjiang populations.

However, all those ancient DNA studies of Xinjiang were confined to the northern and eastern parts of this region. Considering the large geographic range and diverse ecosystems of Xinjiang, such studies in western Xinjiang are in great need to gain a more comprehensive understanding of the prehistoric demography of Xinjiang populations. In recent decades, a number of cultural remains and archaeological sites in western Xinjiang, showing the traits that are characteristic of the middle and late Bronze Age Eurasian Steppe (Steppe_MLBA) cultures (e.g., Sintashta and Andronovo) (Shao and Zhang, 2019), were investigated. However, the stable isotope analysis of the Bronze Age Xiabandi (XBD) population provided direct evidence of wheat and millet consumption in the eastern part of the Pamir Plateau (Zhang et al., 2016), suggesting that the possible East–West cultural interactions and communications in westernmost Xinjiang can be dated to 1,500 BC. A craniometry study on individuals from the Liushui cemetery (∼2,950 BP) in western Xinjiang also showed that the population was already admixed between the East and West Eurasians but with the majority inherited from the former (Zhang et al., 2011). The above research presented a complex and confusing scenario of western Xinjiang. More genetic studies on ancient populations in this region will undoubtedly provide important clues to the issue.

In this study, we collected 15 ancient samples from the XBD cemetery, the earliest archaeological site excavated in western Xinjiang to the best of our knowledge. We then enriched and sequenced the complete mitochondrial genomes of the XBD individuals through designed target probes. By comparing the mitochondrial DNA (mtDNA) of the XBD individuals with that of ancient and extant Eurasians, we explored the early population movement in western Xinjiang.

## MATERIALS AND METHODS

## Archaeological Background, Sampling, and Sequencing

The XBD cemetery is located in the westernmost region of Xinjiang, adjacent to the eastern edge of the Pamir Plateau (**Figure 1A**). This region lies at the intersection of the southern and northern branches of the historical "Silk Road," making it an important melting place for populations from East and Central Asia, as well as those from the Eurasia steppe. The XBD cemetery was investigated by Xinjiang Cultural Relics and Archaeology Institute in 2003 and 2004. The whole cemetery can be divided into three phases, the earliest of which was dated to the Bronze Age (3,500–3,300 BP), and the remaining two phages were dated to Han-Tang (∼2,200–1,300 BP) and Ming-Qing dynasties (∼600–300 BP) (Wu, 2012). The excavations of the jars with contracting neck, the bowls, the trumpet-shaped earrings, as well as the wide band-shaped bracelets in the first phase suggest that the XBD cemetery belongs to the Andronovo culture (**Figure 1B**; Wu, 2012). The cemetery contained 92 burials from the Bronze Age, but only 27 human skeletons were excavated. We selected 15 well-preserved skulls and sampled the intact and sound teeth for genetic research (**Table 1**; **Supplementary Table 1A**). The permission for the use of the 15 Bronze Age samples of the
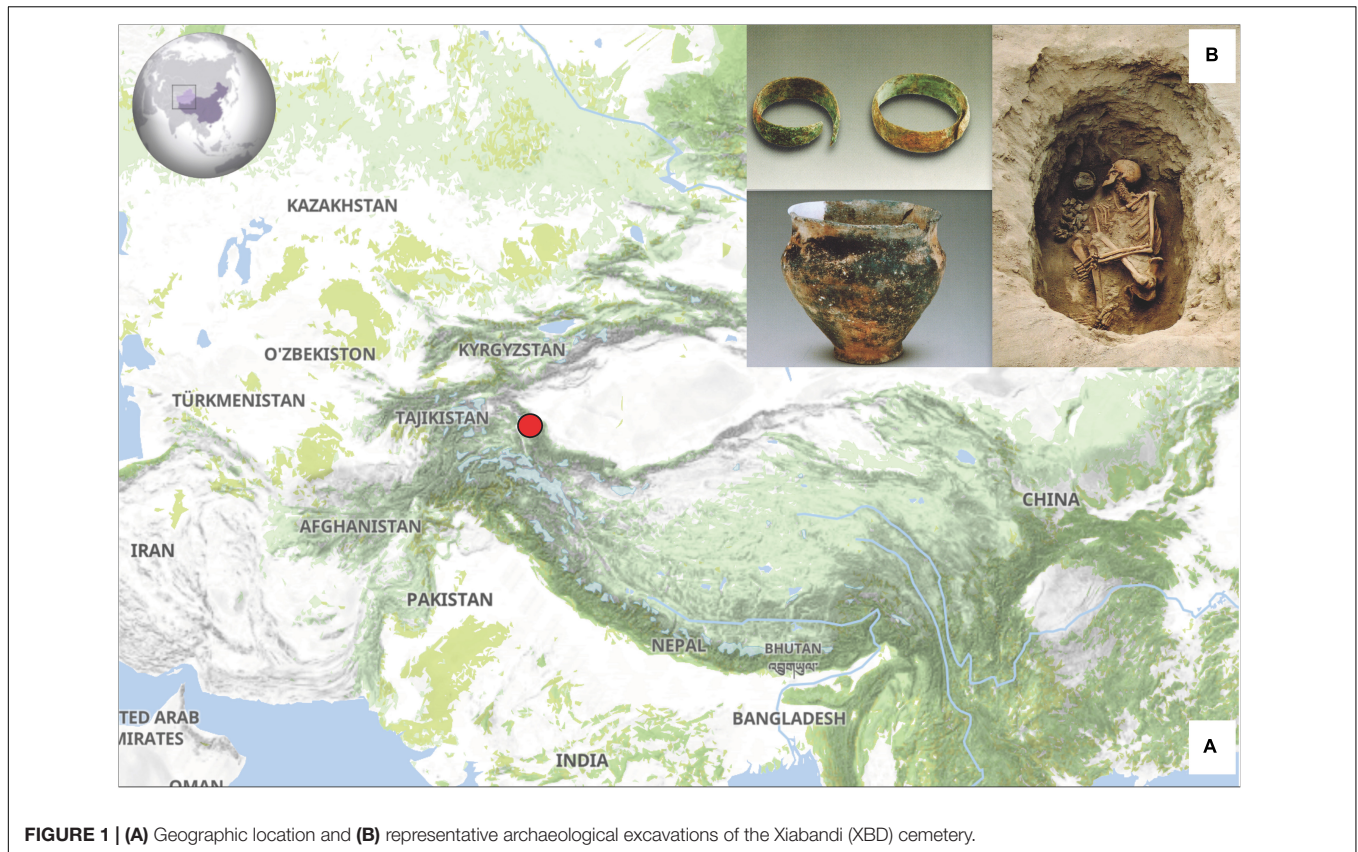
FIGURE 1 | (A) Geographic location and (B) representative archaeological excavations of the Xiabandi (XBD) cemetery.

TABLE 1 | Summary of XBD individuals included in this study.

| Sample ID | Biological sex | Archaeological dating (BP) | Archaeological culture | Mean coverage | Contamination (%) | Haplogroup |
|-----------|----------------|----------------------------|------------------------|---------------|-------------------|------------|
| XBD-M18 | Female | 3,500–3,300 | Andronovo | 1,578 | 1.3 | R1b |
| XBD-40 | Male | 3,500–3,300 | Andronovo | 1,622 | 5.8 | H6a1a |
| XBD-M24 | Male | 3,500–3,300 | Andronovo | 1,143 | 1.6 | H5b |
| XBD-M29 | Male | 3,500–3,300 | Andronovo | 4,130 | 0.6 | U4a1 |
| XBD-M10 | Male | 3,500–3,300 | Andronovo | 187 | 2.8 | U4c1 |
| XBD-M9 | Female | 3,500–3,300 | Andronovo | 3,410 | 3.4 | H6a1a |
| XBD-M36 | Female | 3,500–3,300 | Andronovo | 1,881 | 1.9 | U1a1c1 |
| XBD-36 | Female | 3,500–3,300 | Andronovo | 2,023 | 1.4 | R1b1 |
| XBD-M45 | Male | 3,500–3,300 | Andronovo | 897 | 2.6 | H11b |
| XBD-M46 | Female | 3,500–3,300 | Andronovo | 383 | 2 | HV14 |
| XBD-37 | Female | 3,500–3,300 | Andronovo | 1,308 | 4.8 | U2e2a1d |
| XBD-M48 | Male | 3,500–3,300 | Andronovo | 576 | 2.1 | U2e3 |
| XBD-M17 | Male | 3,500–3,300 | Andronovo | 812 | 2.5 | I4a |
| XBD-M14 | Male | 3,500–3,300 | Andronovo | 1,084 | 4.1 | U2e1 |
| XBD-38 | Male | 3,500–3,300 | Andronovo | 581 | 1.1 | T2a1b1 |

XBD, Xiabandi.

XBD cemetery was obtained from Xinjiang Cultural Relics and Archaeology Institute.

DNA was extracted from teeth powder (∼50 mg) with the method previously described (Ning et al., 2016). The libraries were prepared with the NEBNext Ultra DNA Library preparation kit (New England Biolabs, United Kingdom) following the manufacturer's protocol but with a 1:20 dilution of the adapter during the ligation step. The quality and concentration of the libraries were determined on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, United States). Subsequently, targeted enrichment of the mtDNA was conducted with the MitoCap[TM] kit (MyGenostics, Beijing, China). Sequencing was carried out on an Illumina HiSeq 2000 platform at Novogene Inc. (Beijing, China).

## Sequence Mapping and Mitochondrial DNA Haplogroup Determination

Raw data was processed using EAGER v1.92.50 with default parameters, a pipeline specially designed to deal with ancient DNA data (Peltzer et al., 2016). Quality assessment was performed with FastQC software (Andrews, 2010). The adapters were trimmed with AdapterRemoval v2.2.0 with a minimum overlap of 1 bp and base quality larger than 20 (Schubert et al., 2016). Reads shorter than 30 bp were disregarded. BWA v0.7.12 was used to align the reads to the Revised Cambridge Reference Sequences (rCRS) with seed disabled (-l 2000). The duplicate reads were removed by the DeDup v0.12.1 (Peltzer et al., 2016). Ancient DNA deamination rates were calculated with MapDamage v2.0 (Jonsson et al., 2013). Single-nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) were called using SNVer-0.5.2 (Wei et al., 2011) and were checked by visual inspection. We used trimmed 10 bp at both 3′ and 5′ ends with TrimBam function in the BamUtils v1.0.13[1] to minimize the bias caused by ancient DNA deamination. The mitochondrial haplogroups were determined with Haplogrep2 (Weissensteiner et al., 2016) according to PhyloTree build 17 (Van Oven, 2015).

## Analysis of Xiabandi Mitochondrial DNA Genomes

Haplogroup frequencies were estimated by simple counting. A principal component analysis (PCA) based on the frequencies of sub-haplogroups was performed with the R libraries "factoextra", "FactoMineR", and "ggplot2".

The coalescence time of each lineage was estimated using the ρ statistic-based method and the maximum likelihood (ML) method implemented in PAML software v4.9g (Yang, 2007) with the Soares rate for complete mitochondrial genomes (Soares et al., 2009). For the ρ-based method, the corresponding standard deviation (SD) was calculated following published methods (Saillard et al., 2000). With the knowledge of coalescence time of each haplogroup estimated by contemporary samples, a Bayesian method implemented in BEAST software v1.8.0 was used to infer the time of XBD samples (Drummond and Rambaut, 2007).

## RESULTS

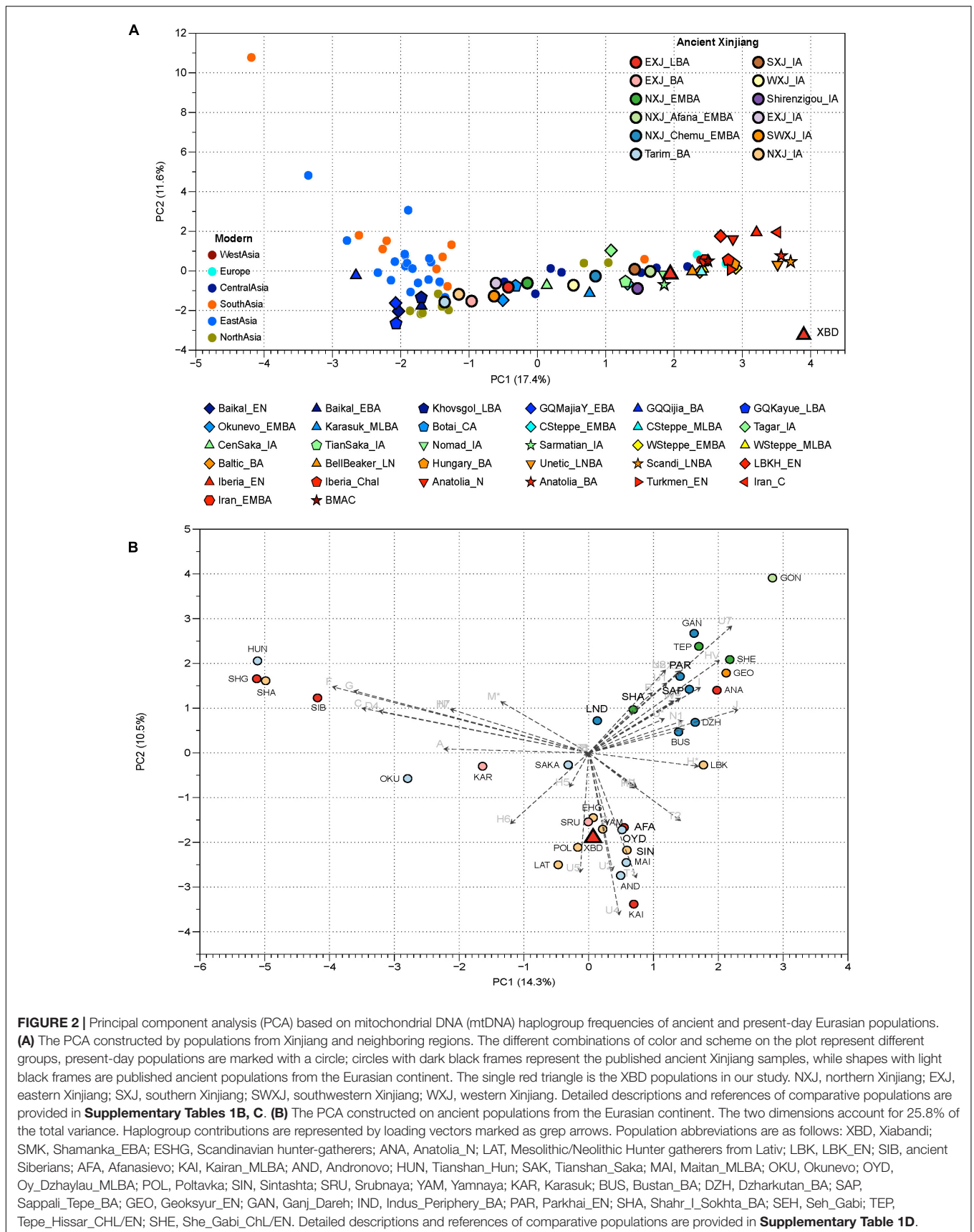## Mitochondrial DNA Authentication and Contamination Assessment

Strict contamination precautions for ancient DNA were taken, and wet lab works were carried out in a dedicated clean room facility specially designed for ancient DNA studies at Jilin University. All samples showed a short fragment length (55–90 base pairs) and postmortem chemical modifications at 3′ and 5′ ends that are expected for ancient DNA (Dabney et al., 2013). The contamination rates of those samples were further quantitatively evaluated with contamMix v1.0-10 (Fu et al., 2013). As a result, we obtained a low level of modern human DNA contaminations

(<5%) with the exception of one individual to be 5.8% (**Table 1**), which restricted our analysis to the deaminated reads to rule out modern human contamination. By applying the above criteria, we confirmed the authentication of our ancient data.

## Major Bronze Age Steppe Pastoralist Origin of the Xiabandi Mitochondrial Haplogroups

We obtained complete mitochondrial sequences to an average coverage of 187- to 4,130-fold across all 15 individuals sequenced in this study. A total of 14 different mitochondrial haplogroups belonging to five macro-haplogroups, such as U, H, T, R, and I, were observed (**Table 1**; **Supplementary Table 1A**). All these haplogroups commonly appear in present-day populations from Europe, Central Asia, and Central/Western steppe, and are uniformly considered to be of West Eurasian origin. Two out of the 14 haplogroups, namely, HV14 and U1a1c1, are prevalent in the extant Central and Western Asians (Palanichamy et al., 2015; Narasimhan et al., 2019; Shamoon-Pour et al., 2019). Haplogroup HV14 was present in two ancient individuals from Central Asia, one (3,000–2,200 BCE) from Turkmenistan and the other (2,100–1,800 BCE) from Uzbekistan (Narasimhan et al., 2019). Similarly, U1a1c1 was found in a historical individual (680–649 CE) from the Pontic steppe (Narasimhan et al., 2019) and the Bronze Age individuals from Iran (3,328–3,022 BCE) and Turkmenistan (2,500–1,700 BCE), who were associated with the bactria-margiana archaeological complex (BMAC). These results suggest that haplogroup HV14 and U1a1c1 are probably Central or Western Asian origin. The remaining 12 haplogroups (I4a, H6a1a, H5b, H11b, R1b, R1b1, T2a1b1, U2e1, U2e2a1d, U2e3, U4a1, U4c1), however, were detected in the Bronze Age steppe pastoralists, the ancient Xinjiang groups, and the prehistoric populations in Europe. For example, haplogroup U2e, the most abundant type in XBD (20%), was found in high frequency in the Sintashta (11.6%) and Andronovo (14.3%) populations. Haplogroup U4a1, which had a high frequency in the Andronovo population (19%), was also observed in one individual associated with the Afanasievo culture (Allentoft et al., 2015). Beyond their wide distributions in the Bronze Age steppe pastoralists, several haplogroups were detected in some pre-Bronze Age hunter-gatherers from the central steppe as well. For example, haplogroup R1b was identified in an Upper Paleolithic individual from the left bank of the Yenisei River dated to around 14,000 BP. In addition, haplogroups U4a1, R1b1, and U2e3 were observed in the Botai culture from northern Kazakhstan and in Eastern Europe hunter-gatherer (Mathieson et al., 2015; Fu et al., 2016; Mittnik et al., 2018). Notably, haplogroups I4a, R1b1, and U2e2a1d were found in individuals who were associated with the BMAC culture and dated to the beginning of this culture 451 in Central Asia. These earlier individuals shared the substratum with the BMAC group but harbored additional Bronze Age steppe pastoralist ancestry than the main BMAC group as evidenced by the autosomal data (Narasimhan et al., 2019). Genetic frequency-based principal component analysis (PCA) agrees with what we have observed in the mitochondrial haplogroup distributions that the XBD falls within the western Eurasian cluster (right) formed

---

[1]https://github.com/statgen/bamUtil

**FIGURE 2 |** Principal component analysis (PCA) based on mitochondrial DNA (mtDNA) haplogroup frequencies of ancient and present-day Eurasian populations. **(A)** The PCA constructed by populations from Xinjiang and neighboring regions. The different combinations of color and scheme on the plot represent different groups, present-day populations are marked with a circle; circles with dark black frames represent the published ancient Xinjiang samples, while shapes with light black frames are published ancient populations from the Eurasian continent. The single red triangle is the XBD populations in our study. NXJ, northern Xinjiang; EXJ, eastern Xinjiang; SXJ, southern Xinjiang; SWXJ, southwestern Xinjiang; WXJ, western Xinjiang. Detailed descriptions and references of comparative populations are provided in **Supplementary Tables 1B, C**. **(B)** The PCA constructed on ancient populations from the Eurasian continent. The two dimensions account for 25.8% of the total variance. Haplogroup contributions are represented by loading vectors marked as grep arrows. Population abbreviations are as follows: XBD, Xiabandi; SMK, Shamanka_EBA; ESHG, Scandinavian hunter-gatherers; ANA, Anatolia_N; LAT, Mesolithic/Neolithic Hunter gatherers from Lativ; LBK, LBK_EN; SIB, ancient Siberians; AFA, Afanasievo; KAI, Kairan_MLBA; AND, Andronovo; HUN, Tianshan_Hun; SAK, Tianshan_Saka; MAI, Maitan_MLBA; OKU, Okunevo; OYD, Oy_Dzhaylau_MLBA; POL, Poltavka; SIN, Sintashta; SRU, Srubnaya; YAM, Yamnaya; KAR, Karasuk; BUS, Bustan_BA; DZH, Dzharkutan_BA; SAP, Sappali_Tepe_BA; GEO, Geoksyur_EN; GAN, Ganj_Dareh; IND, Indus_Periphery_BA; PAR, Parkhai_EN; SHA, Shahr_I_Sokhta_BA; SEH, Seh_Gabi; TEP, Tepe_Hissar_CHL/EN; SHE, She_Gabi_ChL/EN. Detailed descriptions and references of comparative populations are provided in **Supplementary Table 1D**.

by the ancient nomads, WSteppe_EMBA and WSteppe_MLBA, represented by the Yamnaya and Andronovo, respectively (**Figure 2A**; **Supplementary Table 1F**). When compared to the other ancient populations from Xinjiang, we found that the XBD clustered with the NXJ_Afana_EMBA and Shirenzigou_IA, both of which were previously proven to share significant genetic affinity with the Bronze Age steppe pastoralists (Ning et al., 2019; Wang W. et al., 2021). In a finer scale PCA plot, the XBD also clustered with multiple WSH groups but shifted toward the South/Central Asian populations (top right) slightly (**Figure 2B**; **Supplementary Table 1E**), documenting that the majority of XBD mitochondrial haplogroups (12/14) can trace their origin from the Eurasia steppe pastoralist while the minority (2/14) from West or Central Asia.

# Expansion of the Bronze Age Steppe Pastoralists as a Dynamic Process to Form the Genetic Landscape of Xiabandi Individuals

We used 540 present-day mitochondrial sequences obtained from PhyloTree database (van Oven and Kayser, 2009) who were genetically close to XBD individuals to construct the mtDNA phylogeny (**Supplementary Table 2**). Coalescence times of 14 mtDNA haplogroups related to XBD samples were estimated employing the ρ-based and the ML methods. The estimates obtained by both methods showed consistency (**Table 2**), suggesting the reliability of our estimates. Out of the 14 haplogroups, seven (U2e2a1d, I4a, U1a1c1, U4a1, U4c1, H6a1a,

**TABLE 2 |** TMRCA of XBD mtDNA haplogroups estimated from modern mtDNA genomes.

| Haplogroup/sample | Coalescence time (kya) | | |
|---|---|---|---|
| | ρ-based method[a] | ML method | Bayesian method[b] |
| I4a | 6.89 ± 0.93 | 5.88 (4.05, 7.73) | – |
| S02914 | – | – | 5.25 (0.11, 7.71) |
| H6a1a | 6.34 ± 0.66 | 5.94 (4.56, 7.33) | – |
| S02912 | – | – | 5.33 (0.54, 7.38) |
| S04760 | – | – | 4.63 (1.28, 7.33) |
| H5b | 9.21 ± 1.71 | 9.20 (6.29, 12.16) | – |
| H5b + G263A + G7897A | 7.22 ± 2.67 | 6.23 (2.13, 10.44) | – |
| R1769 | – | – | 4.38 (0.51, 8.62) |
| H11b | 10.19 ± 2.81 | 9.26 (3.60, 15.11) | – |
| R1772 | – | – | 4.61 (0.00, 10.43) |
| HV14 | 13.70 ± 4.36 | 15.36 (5.43, 25.81) | – |
| R1773 | – | – | 14.66 (1.92, 22.25) |
| R1b + A9894G + C10160T + T16224C + G16390A | 13.37 ± 4.13 | 14.18 (2.40, 26.74) | – |
| S02915 | – | – | 5.91 (0.00, 17.68) |
| R1b1 | 19.79 ± 3.43 | 20.95 (12.51, 29.72) | – |
| R1b1 + T16189C! | – | – | 17.83 (7.81, 29.03) |
| S04757 | – | – | 12.27 (1.03, 26.14) |
| T2a1b1 | 12.22 ± 2.97 | 10.04 (7.12, 13.02) | – |
| S04759 | – | – | 3.77 (0.00, 7.65) |
| U1a1c1 | 13.91 ± 2.59 | 13.75 (8.18, 19.50) | – |
| R1767 | – | – | 11.29 (3.61, 17.59) |
| U2e1 | 15.21 ± 2.08 | 15.85 (12.36, 19.41) | – |
| U2e1 + C2526T + G12618A | 7.89 ± 2.30 | 9.20 (3.09, 15.54) | – |
| S02913 | – | – | 6.08 (1.31, 11.40) |
| U2e2a1d | 4.16 ± 1.08 | 4.47 (1.87, 7.12) | – |
| S04758 | – | – | 3.10 (0.00, 5.23) |
| U2e3 | 11.79 ± 3.65 | 11.37 (0.78, 22.63) | – |
| U2e3 + C16400T | 11.99 ± 3.90 | 7.71 (1.54, 14.11) | – |
| U2e3 + C16400T + A3892G + T6050C + T10463C + T16017C + T16224C | – | – | 3.23 (0.15, 9.59) |
| R1774 | – | – | 2.09 (0.00, 8.45) |
| U4a1 | 8.93 ± 1.47 | 9.44 (6.53, 12.41) | – |
| U4a1 + C7868T + A13773G | – | – | 6.00 (2.70, 10.75) |
| R1770 | – | – | 4.07 (0.02, 9.85) |
| U4c1 | 6.09 ± 0.88 | 5.64 (3.81, 7.49) | – |
| U4c1 + A827G | – | – | 5.56 (2.94, 7.68) |
| R0341 | – | – | 1.97 (0.00, 7.16) |

[a] We excluded Xiabandi and other ancient samples when using ρ-based method and ML method to estimate the coalescence time of each haplogroup.
[b] We used a Bayesian method to infer the time of Xiabandi samples with the knowledge of coalescence time of each haplogroup estimated by contemporary samples.
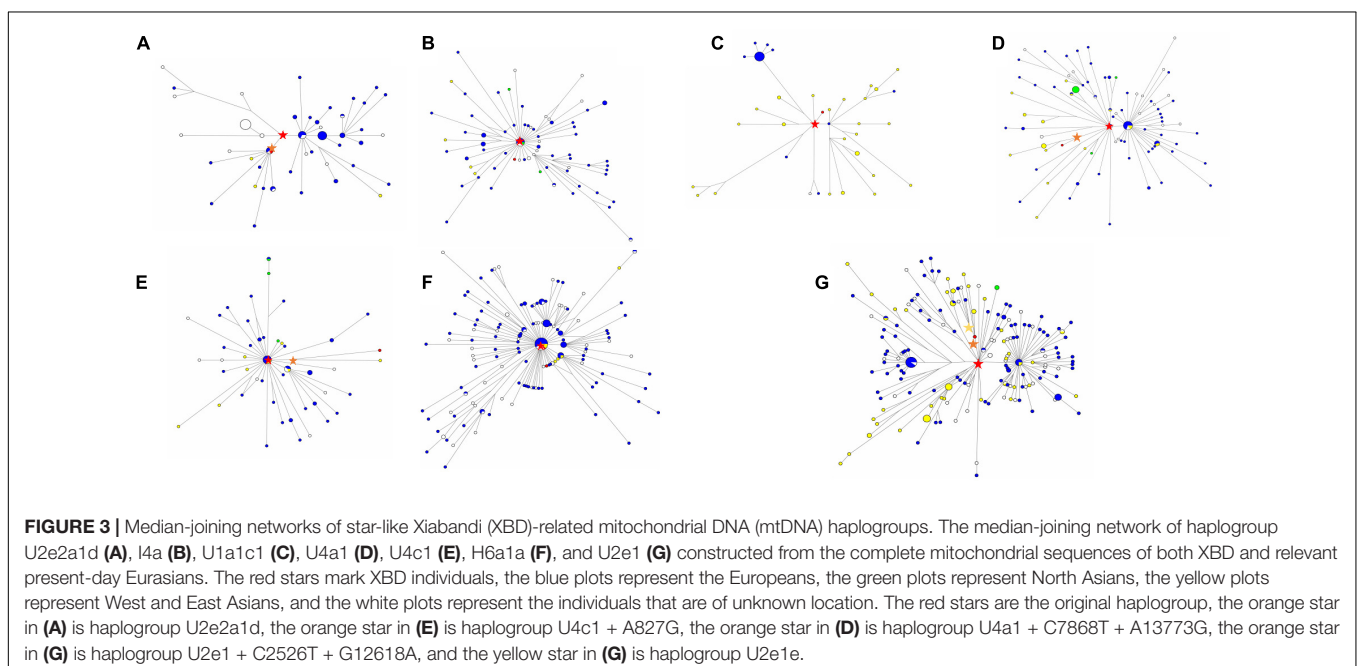ML, maximum likelihood; mtDNA, mitochondrial DNA; XBD, Xiabandi.

and U2e1) showed a star-like phylogeny of their ancestral node, indicating strong population expansions. Among the seven star-like lineages (**Figure 3**), four (I4a, H6a1a, U2e2a1d, and U4c1) were estimated of rather time to most recent common ancestor (TMRCA) of <6,000 BP with the most recent expansion lineage (U2e2a1d) estimated at approximately 4,470 BP (**Table 2**). This time is within the range of the presence of Early Bronze Age steppe pastoralists represented by the Yamnaya culture (3,300–2,700 BCE) in the Pontic steppe and the Afanasievo culture (3,300–2,500 BCE) in the Altai Mountains and fits well with the onset of the Sintashta culture (2,200–1,800 BCE). The Sintashta culture first emerged in the Urals at around 2,200 BC with multiple technological innovations, such as the earliest known chariots and training horses (Kristiansen and Larsson, 2005), and gave rise to the Andronovo culture (1,500–1,700 BC) (Kuznetsov, 2006; Hanks et al., 2007; Allentoft et al., 2015). Those innovations together with the populations quickly spread across much of the Eurasia Steppe (Narasimhan et al., 2019; Jeong et al., 2020). The genetic observations here, as well as the archaeological evidences, suggest that the XBD population originated in a large extent from the middle and late Bronze Age steppe pastoralists, who expanded to the western Xinjiang carrying their technologies along. However, two haplogroups (HV14 and U1a1c1) with Western or Central Asian origin were estimated of rather ancient TMRCA (14,660 and 11,290 BP, respectively) (**Table 2**), suggesting Western or Central Asian to be the source of these two haplogroups. A scenario to explain this phenomenon is that the Bronze Age steppe pastoralists expanded from the western and central steppe southward into Central Asia and admixed in a small scale with the indigenous populations there to form the ancestor of the XBD population, who then migrated eastward over the Pamir Plateau into western Xinjiang. This scenario is consistent with the recent ancient genomic study that the Bronze Age steppe pastoralists only marginally admixed with the indigenous population in

Central Asia they met and moved farther southward into South Asia and admixed extensively with the local populations there (Narasimhan et al., 2019).

## DISCUSSION

The prehistory of Xinjiang is of considerable interest given its special geographic location in connecting the East and the West Eurasians. Multiple genetic studies showed that since the Bronze Age, the populations in Xinjiang had exhibited high genetic diversity and extensive admixture with various populations (Yao et al., 2004; Zhang et al., 2010). The admixture dating analysis based on linkage disequilibrium for the present-day populations in Xinjiang suggested multiple waves of admixture events (Zhang et al., 2010; Shan et al., 2014a,b; Feng et al., 2017). However, tracing the population prehistory with present-day individuals is prone to be distorted by recent admixture events, which is especially the case for Xinjiang populations because the opening of the well-known "Silk Road" made the gene flow among different populations in this region even more frequent. Ancient DNA studies in this region had shown that populations in eastern Xinjiang were already admixed between the East and the West Eurasians as early as the Second Millennium BCE (Li et al., 2010; Wang W. et al., 2021). Population genetic history of western Xinjiang, however, is still largely unknown. By analyzing the XBD mitochondrial genomes, we show here that XBD was genetically admixed from the middle and late Bronze Age steppe pastoralists and the indigenous populations in Central Asia, who probably migrated into Xinjiang through the Pamir Plateau.

The discovery of the Tocharian manuscript from the northern rim of the Tarim Basin and the Indo-Iranian manuscripts from the southern edge provides direct evidence for the dispersal of Indo-European languages into the region (Di Cosmo, 2002). It is now a general consensus among the linguistics that the dispersal of both languages is related to the Bronze Age Steppe herders



**FIGURE 3 |** Median-joining networks of star-like Xiabandi (XBD)-related mitochondrial DNA (mtDNA) haplogroups. The median-joining network of haplogroup U2e2a1d **(A)**, I4a **(B)**, U1a1c1 **(C)**, U4a1 **(D)**, U4c1 **(E)**, H6a1a **(F)**, and U2e1 **(G)** constructed from the complete mitochondrial sequences of both XBD and relevant present-day Eurasians. The red stars mark XBD individuals, the blue plots represent the Europeans, the green plots represent North Asians, the yellow plots represent West and East Asians, and the white plots represent the individuals that are of unknown location. The red stars are the original haplogroup, the orange star in **(A)** is haplogroup U2e2a1d, the orange star in **(E)** is haplogroup U4c1 + A827G, the orange star in **(D)** is haplogroup U4a1 + C7868T + A13773G, the orange star in **(G)** is haplogroup U2e1 + C2526T + G12618A, and the yellow star in **(G)** is haplogroup U2e1e.

(Walter, 1998). The Tocharians may have moved eastward earlier than the Indo-Iranians. The Tocharians are likely to be closely associated with the Afanasievo culture in the Altai Mountains who were a successor of the Yamnaya culture in the Pontic–Caspian Steppe. The middle and late Bronze Age steppe pastoralists, such as the Sintashta, Andronovo, and Srubnaya, are believed to be associated with the dispersal of Indo-Iranian languages (Lamberg-Karlovsky, 2002). The Iron Age individuals from northeastern Xinjiang were proved by autosomal DNA to harbor the Yamnaya/Afanasievo ancestry instead of the Steppe_MLBA, providing a strong genetic link of the "steppe hypothesis" over the "oasis hypothesis" and genetic support for the introduction of the Tocharian languages into Xinjiang (Ning et al., 2019). Our study here suggests a different genetic profile of totally west Eurasian origin, and provides a genetic link for the existence of Indo-Iranian languages in western Xinjiang at least 3,300 years ago.

## CONCLUSION

Taken together, the systematic mtDNA analysis on ancient samples from the westernmost part of Xinjiang provides us a unique opportunity to investigate the population origin of Xinjiang with a broader geography. We find that the 15 XBD individuals fall within the range of the ancient western Eurasian variation, and the formation of the ancestry legacy of XBD is related to the expansion of the middle and late Bronze Age steppe herders who might speak Indo-Iranian languages and admixed with the indigenous populations in the West or Central Asia during their expansion. Additionally, integrating the archaeological and genetic evidences in this study, the existence of the Andronovo culture in western Xinjiang involved not only the dispersal of ideas but also population movement. We recognize that such study on samples from a broader region and time sequences is required to obtain a more comprehensive understanding of the population prehistory of Xinjiang.

## AUTHOR CONTRIBUTIONS

YC, LJ, and SG conceived and supervised the study. CN, YZ, YX, and CL performed the laboratory work. YW and DW provided archaeological materials and associated information. CN, H-XZ, FZ, and SW analyzed the data. CN, YC, FZ, SG, and H-XZ wrote the manuscript with the input from all co-authors.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.740167/full#supplementary-material

## REFERENCES

Allentoft, M. E., Sikora, M., Sjogren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172. doi: 10.1038/nature14507

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., et al. (2013). Complete mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort DNA fragments. *PNAS* 110, 15758–15763. doi: 10.1073/pnas.1314445110

Damgaard, P. B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., et al. (2018a). 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374. doi: 10.1038/s41586-018-0094-2

Damgaard, P. B., Martiniano, R., Kamm, J., Moreno-Mayar, J. V., Kroonen, G., Peyrot, M., et al. (2018b). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360:eaar7711. doi: 10.1126/science.aar7711

Di Cosmo, N. (2002). *The Tarim Mummies: ancient China and the mystery of the earliest peoples from the West*. London: Thames & Hudson, 279–281.

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Feng, Q., Lu, Y., Ni, X., Yuan, K., Yang, Y., Yang, X., et al. (2017). Genetic History of Xinjiang's Uyghurs Suggests Bronze Age Multiple-Way Contacts in Eurasia. *Mol. Biol. Evol.* 34, 2572–2582. doi: 10.1093/molbev/msx177

Frachetti, M. D. (2009). *Pastoralist Landscapes And Social Interaction In Bronze Age Eurasia*. Berkeley, California: University of California Press, doi: 10.1525/9780520942691

Fu, L., Yuan, J., and Li, S. (2009). The origin and character of domesticated animals during Neolithic Age in Gansu–Qinghai region. *Archaeology* 5, 80–86.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23, 553–559. doi: 10.1016/j.cub.2013.02.044

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., et al. (2016). The genetic history of Ice Age Europe. *Nature* 534, 200–205. doi: 10.1038/nature17993

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

Hanks, B. K., Epimakhov, A. V., and Renfrew, A. C. (2007). Towards a refined chronology for the Bronze Age of the southern Urals, Russia. *Antiquity* 81, 353–367. doi: 10.1017/S0003598X00095235

Jeong, C., Wang, K., Wilkin, S., Taylor, W. T. T., Miller, B. K., Bemmann, J. H., et al. (2020). A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell* 183, 890–904.e29. doi: 10.1016/j.cell.2020.10.015

Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A. S., Taylor, W. T. T., Hagan, R. W., et al. (2018). 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. doi: 10.1073/pnas.1813608115

Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684. doi: 10.1093/bioinformatics/btt193

Kristiansen, K., and Larsson, T. B. (2005). *The Rise Of Bronze Age Society : travels, Transmissions And Transformations*. Cambridge; New York: Cambridge University Press.

Kuznetsov, P. F. (2006). The emergence of Bronze Age chariots in eastern Europe. *Antiquity* 80, 638–645. doi: 10.1017/S0003598X00094096

Lamberg-Karlovsky, C. C. (2002). Archaeology and language: the indo-iranians. *Current Anthropology* 43, 63–88. doi: 10.1086/324130

Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W., et al. (2010). Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol.* 8:15. doi: 10.1186/1741-7007-8-15

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503. doi: 10.1038/nature16152

Mittnik, A., Wang, C. C., Pfrengle, S., Daubaras, M., Zarina, G., Hallgren, F., et al. (2018). The genetic prehistory of the Baltic Sea region. *Nat. Commun.* 9:442. doi: 10.1038/s41467-018-02825-9

Narasimhan, V. M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., et al. (2019). The formation of human populations in South and Central Asia. *Science* 365:eaat7487. doi: 10.1126/science.aat7487

Ning, C., Gao, S., Deng, B., Zheng, H., Wei, D., Lv, H., et al. (2016). Ancient mitochondrial genome reveals trace of prehistoric migration in the east Pamir by pastoralists. *J. Hum. Genet.* 61, 103–108. doi: 10.1038/jhg.2015.128

Ning, C., Wang, C. C., Gao, S., Yang, Y., Zhang, X., Wu, X., et al. (2019). Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan. *Curr. Biol.* 29, 2526–2532. doi: 10.1016/j.cub.2019.06.044

Palanichamy, M. G., Mitra, B., Zhang, C. L., Debnath, M., Li, G. M., Wang, H. W., et al. (2015). West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. *Hum. Genet.* 134, 637–647. doi: 10.1007/s00439-015-1547-4

Peltzer, A., Jager, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., et al. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biol.* 17:60. doi: 10.1186/s13059-016-0918-z

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J., and Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* 67, 718–726. doi: 10.1086/303038

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9:88. doi: 10.1186/s13104-016-1900-2

Shamoon-Pour, M., Li, M., and Merriwether, D. A. (2019). Rare human mitochondrial HV lineages spread from the Near East and Caucasus during post-LGM and Neolithic expansions. *Sci. Rep.* 9:14751. doi: 10.1038/s41598-019-48596-1

Shan, W., Ablimit, A., Zhou, W., Zhang, F., Ma, Z., and Zheng, X. (2014a). Genetic polymorphism of 17 Y chromosomal STRs in Kazakh and Uighur populations from Xinjiang, China. *Int. J. Legal Med.* 128, 743–744. doi: 10.1007/s00414-013-0948-y

Shan, W., Ren, Z., Wu, W., Hao, H., Abulimiti, A., Chen, K., et al. (2014b). Maternal and paternal diversity in Xinjiang Kazakh population from China. *Genetika* 50, 1374–1385. doi: 10.7868/s0016675814110149

Shao, H., and Zhang, W. (2019). Review of the Research of Andronovo Cultures in Xinjiang (in Chinese). *Western Reg. Stud.* 2, 113–121. doi: 10.16363/j.cnki.xyyj.2019.02.012

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., et al. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* 84, 740–759. doi: 10.1016/j.ajhg.2009.05.001

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.fsigss.2015.09.155

van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30, E386–E394. doi: 10.1002/humu.20921

Walter, M. N. (1998). *Tokharian Buddhism In Kucha: buddhism Of Indo-European Centum Speakers In Chinese Turkesta Before The 10th Century Ce.* United States: University of Pennsylvania.

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, W., Ding, M., Gardner, J. D., Wang, Y., Miao, B., Guo, W., et al. (2021). Ancient Xinjiang mitogenomes reveal intense admixture with high genetic diversity. *Sci. Adv.* 7:eabd6690. doi: 10.1126/sciadv.abd6690

Wei, Z., Wang, W., Hu, P., Lyon, G. J., and Hakonarson, H. (2011). SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Res.* 39:e132. doi: 10.1093/nar/gkr599

Weissensteiner, H., Pacher, D., Kloss-Brandstatter, 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, W58–W63. doi: 10.1093/nar/gkw233

Wood, F. (2002). *The Silk Road: two thousand years in the heart of Asia.* Berkeley, California: Univ of California Press.

Wu, Y. (2012). On the Bronze Culture of Xiabandi Cemetery in Kashi, Xinjiang (in Chinese). *Western Reg. Stud.* 4, 36–44. doi: 10.16363/j.cnki.xyyj.2012.04.013

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Yao, Y. G., Kong, Q. P., Wang, C. Y., Zhu, C. L., and Zhang, Y. P. (2004). Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol. Biol. Evol.* 21, 2265–2280. doi: 10.1093/molbev/msh238

Zhang, F., Xu, Z., Tan, J., Sun, Y., Xu, B., Li, S., et al. (2010). Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am. J. Phys. Anthropol.* 142, 314–320. doi: 10.1002/ajpa.21237

Zhang, J., Wu, X., Li, L., Jin, L., Li, H., and Tan, J. (2011). Cranial Non-metric Evidence for Population Admixture Between East and West Eurasia in Bronze Age,Southwestern Xinjiang (in Chinese). *Renleixue Xuebao* 30, 379–404. doi: 10.16359/j.cnki.cn11-1963/q.2011.04.006

Zhang, X., Wei, D., Wu, Y., Nie, Y., and Hu, Y. (2016). Carbon and nitrogen stable isotope ratio analysis of Bronze Age humans from the Xiabandi cemetery, Xinjiang, China: implications for cultural interactions between the East and West. *Chin. Sci. Bull.* 61, 3509–3519. doi: 10.1360/n972016-00514

Zheng, H. X., Li, L., Jiang, X. Y., Yan, S., Qin, Z., Wang, X., et al. (2017). MtDNA genomes reveal a relaxation of selective constraints in low-BMI individuals in a Uyghur population. *Hum. Genet.* 136, 1353–1362. doi: 10.1007/s00439-017-1829-0

Zhou, X., Yu, J., Spengler, R. N., Shen, H., Zhao, K., Ge, J., et al. (2020). 5,200-year-old cereal grains from the eastern Altai Mountains redate the trans-Eurasian crop exchange. *Nat. Plants* 6, 78–87. doi: 10.1038/s41477-019-0581-y