

# Current Biology

## Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan

### Highlights

- Iron Age Shirezigou individuals are genetically admixed of east and west Eurasians
- The west Eurasian ancestry is most like Yamnaya related
- The east Eurasian ancestry is more like northeast Asian related
- Strong evidence for the introduction of Indo-European languages into Xinjiang

### Authors

Chao Ning, Chuan-Chao Wang, Shizhu Gao, ..., Jian Ma, Johannes Krause, Yinqiu Cui

### Correspondence

eurasiansteppes@126.com (J.M.), krause@shh.mpg.de (J.K.), cuiyq@jlu.edu.cn (Y.C.)

### In Brief

Ning et al. show that the Iron Age genomes from Xinjiang are highly structured with genetic ancestry related to both east and west Eurasians. The finding of Yamnaya/Afanasievo-related ancestry in this region further provides strong evidence for the dispersal of Indo-European language into Xinjiang.

# Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan

Chao Ning,<sup>1,2,8</sup> Chuan-Chao Wang,<sup>3,2,8</sup> Shizhu Gao,<sup>4</sup> Yang Yang,<sup>4</sup> Xue Zhang,<sup>4</sup> Xiyan Wu,<sup>1</sup> Fan Zhang,<sup>1</sup> Zhongzhi Nie,<sup>5</sup> Yunpeng Tang,<sup>6</sup> Martine Robbeets,<sup>2</sup> Jian Ma,<sup>6,\*</sup> Johannes Krause,<sup>2,\*</sup> and Yinqiu Cui<sup>1,5,7,9,\*</sup>

<sup>1</sup>School of Life Sciences, Jilin University, Changchun 130012, China

<sup>2</sup>Max Planck Institute for the Science of Human History, Jena 07745, Germany

<sup>3</sup>Department of Anthropology & Ethnology, Institute of Anthropology, Xiamen University, Xiamen 361005, China

<sup>4</sup>College of Pharmacia Sciences, Jilin University, Changchun 130021, China

<sup>5</sup>Research Center for Chinese Frontier Archaeology, Jilin University, Changchun 130012, China

<sup>6</sup>School of Cultural Heritage, Northwest University, Xi'an 710069, China

<sup>7</sup>Key Laboratory for Evolution of Past Life and Environment in Northeast Asia (Jilin University), Ministry of Education, Changchun 130021, China

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead Contact

\*Correspondence: [eurasiansteppes@126.com](mailto:eurasiansteppes@126.com) (J.M.), [krause@shh.mpg.de](mailto:krause@shh.mpg.de) (J.K.), [cuiyq@jlu.edu.cn](mailto:cuiyq@jlu.edu.cn) (Y.C.)

<https://doi.org/10.1016/j.cub.2019.06.044>

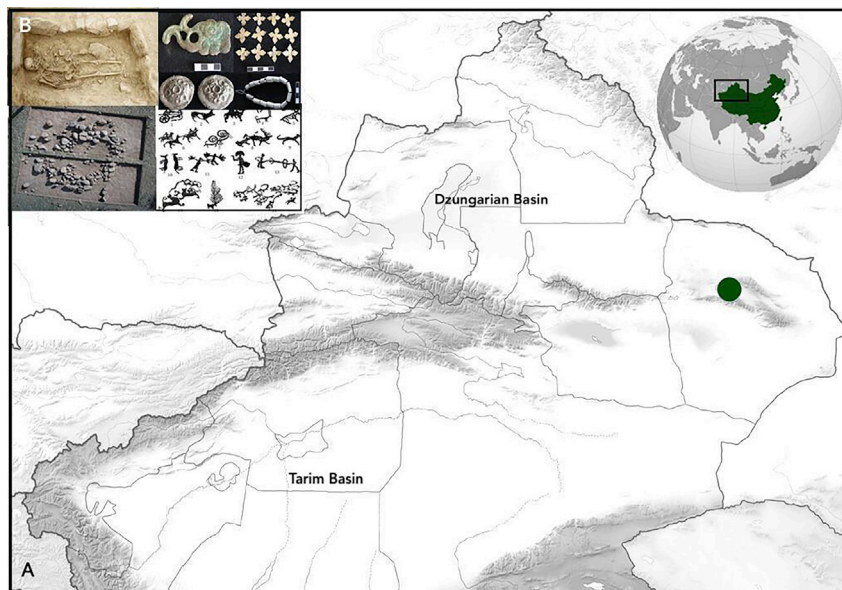
## SUMMARY

Recent studies of early Bronze Age human genomes revealed a massive population expansion by individuals-related to the Yamnaya culture, from the Pontic Caspian steppe into Western and Eastern Eurasia, likely accompanied by the spread of Indo-European languages [1–5]. The south eastern extent of this migration is currently not known. Modern-day human populations from the Xinjiang region in northwestern China show a complex population history, with genetic links to both Eastern and Western Eurasia [6–10]. However, due to the lack of ancient genomic data, it remains unclear which source populations contributed to the Xinjiang population and what was the timing and the number of admixture events. Here, we report the first genome-wide data of 10 ancient individuals from northeastern Xinjiang. They are dated to around 2,200 years ago and were found at the Iron Age Shirengigou site. We find them to be already genetically admixed between Eastern and Western Eurasians. We also find that the majority of the East Eurasian ancestry in the Shirengigou individuals is-related to northeastern Asian populations, while the West Eurasian ancestry is best presented by ~20% to 80% Yamnaya-like ancestry. Our data thus suggest a Western Eurasian steppe origin for at least part of the ancient Xinjiang population. Our findings furthermore support a Yamnaya-related origin for the now extinct Tocharian languages in the Tarim Basin, in southern Xinjiang.

## RESULTS AND DISCUSSION

### Sample Context and Sequencing

Xinjiang, geographically located in northwestern China, has long been one of the major crossroads between eastern and western Eurasia. The Silk Road located in the heart of the Xinjiang documents the intensive connections between the East and the West since at least the Han dynasty (207 BCE–220 CE). Another well-known example is the Tocharian language, which is an extinct branch of the Indo-European language family but was found on the northern edge of the Tarim Basin, in southern Xinjiang during the first millennium CE. The origin of the Tocharian speakers has been long debated. Archaeologists have proposed two hypotheses, the “Bactrian Oasis hypothesis” and the “Steppe hypothesis,” to explain the first appearance of agricultural communities in the region ca. 2,000 BCE [11]. The “Bactrian oasis hypothesis” suggests the oasis-based agriculture may have spread from the Bactria-Margiana Archaeological Complex (BMAC) region with the incursion of nomadic groups particularly from the Andronovo Culture into the Tarim Basin in the early second millennium BCE [12]. In contrast, the “Steppe hypothesis” argues that the initial immigrants to Xinjiang were instead part of the Steppe populations most likely related to the Afanasievo culture from the Altai region north of the Tarim Basin. A number of craniometric studies have been carried out in Xinjiang indicating a degree of mixture between western Eurasian and eastern Eurasian people [11, 13]. Genetic evidence, especially from maternal mtDNA suggested that populations in Xinjiang were genetically admixed between eastern and western Eurasians dating back to at least early Bronze Age [7, 14]. This was further confirmed by a more recent study on genomic autosomal DNA, suggesting that present-day Uyghurs in Xinjiang are genetically derived from various geographic groups, namely, Europeans (25%–37%), South Asians (12%–20%), Siberians (15%–17%), and East Asians (29%–47%) [6]. The Uyghurs show a southwest



**Figure 1. Geographic Location and Archaeological Excavations of the Shirenzigou Site**

(A) Map showing the location of the Shirenzigou site. (B) Archaeological relics excavated from the Shirenzigou site.

to northeast substructure genetically [6]. The paternal lineage enriched in ancient Steppe populations was R1b. But nowadays the haplogroups R1a and R1b have both become characteristic lineages of the Xinjiang populations [15, 16]. However, those previous studies mainly focused on a few present-day populations or a limited number of markers on Y chromosome non-recombination (NRY) and mitochondrial regions, whose resolution is not enough to support either the Bactrian Oasis or the Steppe hypothesis. Therefore, the complicated admixture history between Eastern and Western Eurasian people remains unclear.

To investigate temporal genetic dynamics, we generated genome-wide DNA sequences for 10 human individuals that were collected from the Shirenzigou site, an early Iron Age agro-pastoralist site on the northern slope of eastern Tianshan Mountains, Xinjiang, Northwestern China (Figure 1). This site showed multiple cultural features from the surrounding regions, such as the Yanbulake Culture of the southern slope of the Tianshan Mountains, the Pazyryk Culture from the Altai region as well as the excavated beads imported from the central region of China [17, 18]. Five of the sequenced samples were radiocarbon dated and all fall into the Iron Age (~200–100 BCE) (Table 1). The samples were sequenced to genomic coverage from ~0.04× to 0.75× (Table 1). The sequence reads show DNA damage patterns consistent with ancient DNA (Data S1A) and low contamination estimates for mtDNA (1%–2%) as well as for the X chromosomes in male samples (~0%–1.5% for samples with at least 100 SNPs on X chromosome) (Data S1A). None of the individuals were found to share close kinship (Figure S1).

### Uniparental DNA

The dominant mtDNA lineages of the Shirenzigou people are commonly found in modern and ancient West Eurasian populations, such as U4, U5, and H, while they also have East Eurasian-specific haplogroups A, D4, and G3, preliminarily documenting admixed ancestry from eastern and western Eurasia. The admixture profile is also shown on the paternal Y chromosome side that 4 out of 6 males in Shirenzigou (Figure S2) belong to the

West Eurasian-specific haplogroup R1b (n = 2) and East Eurasian-specific haplogroup Q1a (n = 2) (Table 1), the former is predominant in ancient Yamnaya and nearly 100% in Afanasievo [1, 19], different from the Middle and Late Bronze Age Steppe groups (Steppe\_MLBA) such as Andronovo, Poltavka, Srubnaya, and Sintashta whose Y chromosomal haplogroup is mainly R1a [19–21].

### Autosomal DNA

We called between 27 and 479 thousand SNPs (Table 1; Data S1A) from our ancient samples at 1,240k positions previously described [19]. We merged our newly generated data with published ancient and modern data [2, 3, 22]. We first carried out principal component analysis (PCA) [23] to assess the genetic affinities of the ancient individuals qualitatively by projecting them onto present-day Eurasian variation (Figure 2). We observed a distinct separation between East and West Eurasians. Our ancient Shirenzigou samples and present-day populations from Central Asia and northwestern China form a genetic cline from East to West in the first PC. The distribution of Shirenzigou samples on the cline is relatively scattered with two major clusters, one being closer to modern-day Uyghurs and Kazakhs and the other being closer to recently published ancient Saka and Huns from the Tianshan in Kazakhstan [2]. We then carried out the model-based ADMIXTURE clustering analysis [24] to get a detailed overview of the ancestry composition (Figure 2). Consistent with the PCA, the Shirenzigou samples show a similar East-West admixed genetic profile with present-day populations from Central Asia and northwestern China, for example, Uyghur, Kyrgyz, and Uzbek. But, interestingly, the Shirenzigou samples lack the green component that is enriched in Anatolian and European farmers when compared to the above present-day groups. The East Eurasian component in Shirenzigou looks more related to northern Asians such as Daur and Hezhen than to southern Asians as they do not have the pink component that is enriched in southern Asians (Figure 2).

### Genetic Composition of Iron Age Shirenzigou Individuals

We applied a formal admixture test using  $f_3$  statistics [25] in the form of  $f_3$  (Shirenzigou; X, Y) where X and Y are worldwide populations that might be the genetic sources for the Shirenzigou individuals. We observed the most significant signals of admixture in the Shirenzigou samples when using Yamnaya\_Samara or Srubnaya as the West Eurasian source and some Northern Asians or Koreans as the East Eurasian source (Table S1). We also plotted the outgroup  $f_3$  statistics in the form of  $f_3$  (Mbuti; X, Anatolia\_Neolithic) and  $f_3$  (Mbuti; X, Kostenki14) to visualize

**Table 1. Sample Information and Nuclear Human DNA Screening of All Shirenzigou Individuals**

Sample Name	Skeletal Elements	C14 Dating (cal. BCE)	Mean Coverage	1240k SNPs	mtDNA Haplogroup	Y Haplogroup
M4	petrous bone	380–200	0.0468	42,170	U4'9	–
M8R1	petrous bone	330–200	0.037	26,963	T1a1b	–
M15-1	petrous bone	N/A	0.1404	118,763	I1b	R1b1a1a2
M15-2	petrous bone	N/A	0.1029	98,380	U4	Q1a1a1
X3	petrous bone	N/A	0.1892	164,388	G3b	Q1a1b
M819	tooth	N/A	0.7558	478,956	H15b1	N/A
M820	tooth	285–230	0.2901	246,312	U5b2c	N/A
M012	tooth	N/A	0.1564	139,458	U5a2	R1b1a1a2
M010	tooth	410–350	0.5663	441,036	D4j1b	N/A
F004	petrous bone	370–190	0.321	259,720	A17	N/A

cal. BCE, calibrated years before the common era (given in 2-sigma range); sample without an appropriate radiocarbon dating is marked with N/A; males that are failed in Y chromosome haplogroup assignment are marked with a dash, and females are indicated with N/A (see also [Figure S2](#)).

the allele sharing between population X and Anatolian farmers. As shown in [Figure S3](#), the Steppe\_MLBA populations including Srubnaya, Andronovo, and Sintashta were shifted toward farming populations compared with Yamnaya groups and the Shirenzigou samples. This observation is consistent with ADMIXTURE analysis that Steppe\_MLBA populations have an Anatolian and European farmer-related component that Yamnaya groups and the Shirenzigou individuals do not seem to have [[1–3](#), [19](#), [26](#)]. The analysis consistently suggested Yamnaya-related Steppe populations were the better source in modeling the West Eurasian ancestry in Shirenzigou.

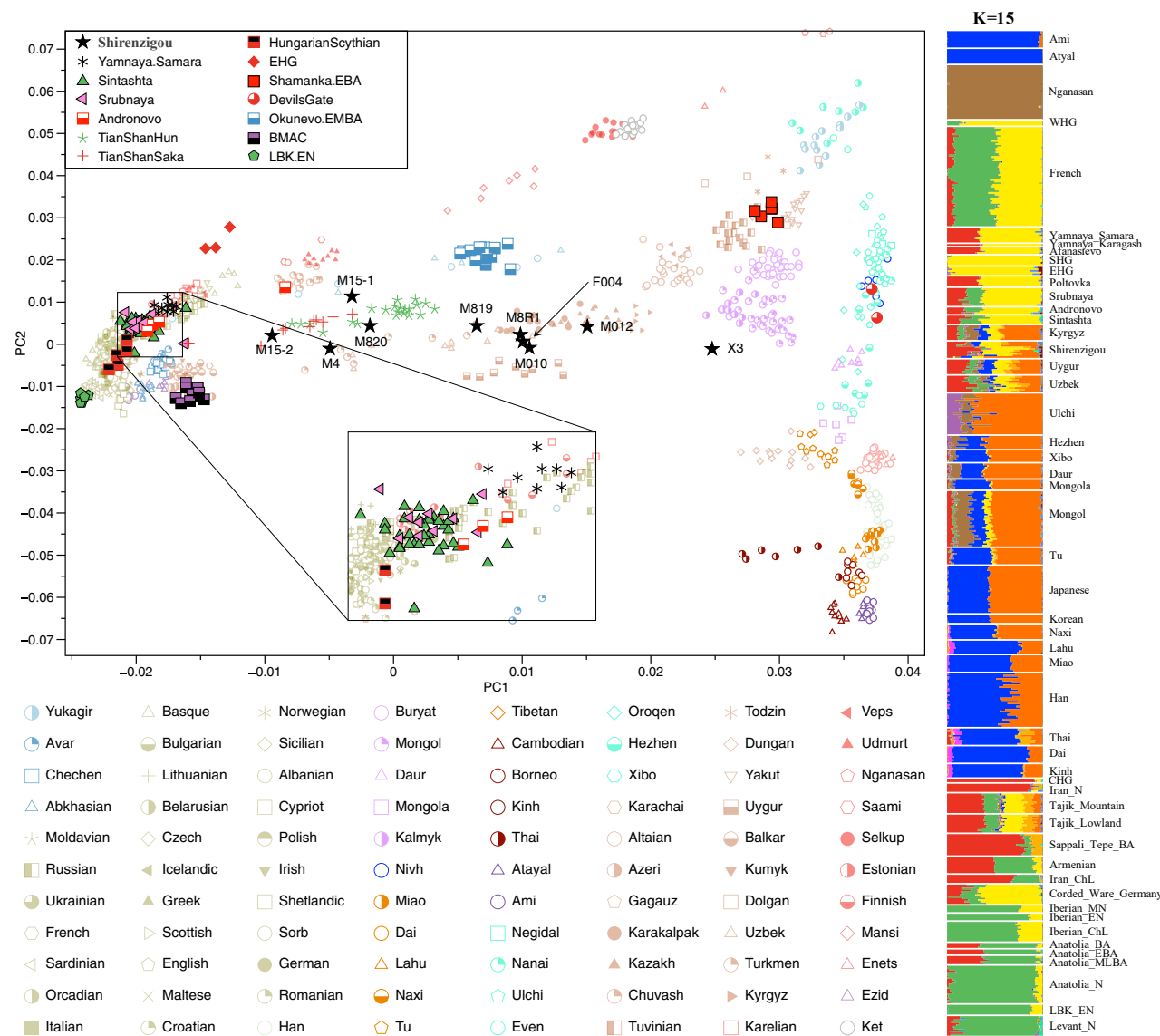
We continued to use *qpAdm* [[4](#), [27](#)] to estimate the admixture proportions in the Shirenzigou samples by using different pairs of source populations, such as Yamnaya\_Samara, Afanasievo, Srubnaya, Andronovo, BMAC culture (Bustan\_BA and Sappali\_Tepe\_BA) and Tianshan\_Hun as the West Eurasian source and Han, Ulchi, Hezhen, Shamanka\_EN as the East Eurasian source. In all cases, Yamnaya, Afanasievo, or Tianshan\_Hun always provide the best model fit for the Shirenzigou individuals, while Srubnaya, Andronovo, Bustan\_BA and Sappali\_Tepe\_BA only work in some cases [[3](#)] ([Tables 2](#) and [S2](#); [Data S2A](#)). The Yamnaya\_Samara or Afanasievo-related ancestry ranges from ~20% to 80% in different Shirenzigou individuals, consistent with the scattered distribution on the East-West cline in the PCA ([Figure 2](#)). In the PCA, ADMIXTURE, outgroup  $f_3$  statistics ([Figure S4](#)), as well as  $f_4$  statistics ([Table S3](#)), we observed the Shirenzigou individuals were closer to the present-day Tungusic and Mongolic-speaking populations in northern Asia than to the populations in central and southern China, suggesting the northern populations might contribute more to the Shirenzigou individuals. Based on this, we then modeled Shirenzigou as a three-way admixture of Yamnaya\_Samara, Ulchi (or Hezhen) and Han to infer the source from the East Eurasia side that contributed to Shirenzigou. We found the Ulchi or Hezhen and Han-related ancestry had a complicated and unevenly distribution in the Shirenzigou samples. The most Shirenzigou individuals derived the majority of their East Eurasian ancestry from Ulchi or Hezhen-related populations, while the following two individuals M820 and M15-2 have more Han related than Ulchi/Hezhen-related ancestry ([Data S2C1](#) and [S2C2](#)).

The origin and dating of the East and West Eurasian admixture in Xinjiang have been long debated. There is a discussion on

when and how many times eastern and western Eurasian populations met and admixed in Xinjiang. Genomic studies estimated the admixture time in present-day Uygur people from Xinjiang ranging from ~150 to 20 generations ago, about 4,200 to 560 years ago assuming a generation time of 28 years [[6](#), [25](#), [28](#), [29](#)]. However, due to the complex history of human populations that may suffer frequent population replacements and also potential gene flows, the dates inferred from present-day populations can only be approximations. By sequencing ancient DNA from archaeological excavations, we can directly infer population changes at various points in time. Here, we report genomic data of 10 Iron Age samples from the Shirenzigou site in the eastern Tianshan. We find the Shirenzigou people were genetically admixed deriving ancestry from both eastern and western Eurasians, tracing the East-West admixture in Xinjiang directly back to at least 2,000 years ago.

The Shirenzigou individuals were found on a genetic cline from East Asians to Western Eurasians showing different East-West admixture proportions beside them being from the same time period and archaeological site. For example, the difference of Yamnaya-related ancestry between individuals from the Shirenzigou site reached up to 80%. One possible explanation is a recent admixture event between groups with Yamnaya-related ancestry and East Asian ancestry. It could furthermore indicate that the region was a corridor and meeting place for Eastern and Western populations or indicate a structured population with various ancestries that occupied the Iron Age Xinjiang region. We also cannot rule out that the genetic heterogeneity had already existed in the western source populations before they entered the eastern Tianshan region to form the Shirenzigou individuals, since the genomic composition of nomadic populations such as Scythian, Hun, and Saka at that period in the region were also highly dynamic [[2](#)].

The existing archaeological evidences suggest that the Shirenzigou site shows typical characteristics of the agro-pastoral Yanbulake Culture from the Bronze-Age Hami Basin located in the southern slope of the East Tianshan Mountains. The animal motifs such as the deer-shaped Griffin in the site also reflects the influences from the Pazyryk Culture from the Altai region. Besides the cultural exchange between the eastern Tianshan Mountains and the Altai region, a special funeral ritual appeared in the Shirenzigou site, whereby the upper torsos were disturbed.



**Figure 2. PCA and ADMIXTURE Analysis for Shirenzigou Samples**

We projected the ancient data in this study as well as the published ancient data onto the first two components defined by a set of selected present-day Eurasians from the “Human Origins” dataset.

See also [STAR Methods](#), [Figures S3](#) and [S4](#), and [Tables S1–S3](#)

This was a common custom in the Neolithic Gansu corridor, in Northwestern China [18]. Furthermore, the chemical composition analysis of ancient glass beads excavated from the site indicated they were imported from the central region of China [17]. The diverse cultural elements observed in the same site provide evidence that different populations once came to this region and admixed with each other to form the genetic structure of the Shirenzigou people, which is well supported by the ancestry profile inferred in this study. We found most Shirenzigou individuals derived a large amount of their ancestry from Ulchi or Hezhen-related populations, which might be associated with ancient nomadic people in northern Asia. The Shirenzigou samples also harbor a Han Chinese-related component, which may be introduced into this region by the farming populations from

surrounding regions, such as Gansu and Qinghai, who also contributed to present-day Han Chinese.

Based on the ADMIXTURE and *qpAdm* analysis, we conclude that the western Eurasian ancestry in the Shirenzigou individuals was most likely related to Yamnaya without significant evidence of European farmer-related gene flow that is present in later Steppe\_MLBA populations [19, 26]. The Yamnaya-related steppe ancestry has been described as a mixture of East-and-Caucasus hunter-gatherers from the Pontic-Caspian steppes, dating to 3,300–2,600 BCE, which eventually spread further to the Altai region in the East in the form of people associated to the Afanasievo Culture. The same population likely migrated to Europe in the West contributing substantially to present-day Europeans along with the spread of some of the

**Table 2. p Values and Ancestry Proportions in Modeling a Two-Way (p = rank 1) Admixture in Shirenzigou Samples Using the Following Set of Outgroups Populations: Dinka, Ust\_Ishim, Kostenki14, Onge, Papuan, Australian, Iran\_N, EHG, and LBK\_EN**

Sample	p = rank 1	Yamnaya_Samara		Han	
		Proportion	SE	Proportion	SE
All	5.996E-02	0.542	0.010	0.458	0.010
M819	8.479E-01	0.574	0.020	0.426	0.020
M820 <sup>a</sup>	2.784E-03	0.731	0.022	0.269	0.022
M15-1	1.248E-01	0.695	0.026	0.305	0.026
M15-2	1.192E-01	0.803	0.029	0.197	0.029
M4	2.682E-01	0.773	0.041	0.227	0.041
M8R1	7.824E-01	0.493	0.047	0.507	0.047
X3	3.336E-01	0.261	0.023	0.739	0.023
M012	2.093E-01	0.400	0.025	0.600	0.025
M010	1.609E-01	0.463	0.020	0.537	0.020
F004	6.563E-01	0.480	0.022	0.520	0.022

<sup>a</sup>Individual failed in *qpAdm* modeling; models with a  $p \geq 0.05$  are considered as a good fit and the failed models have  $p \leq 0.05$

Indo-European languages [1, 4]. In the following millennia, the Altai region was characterized by the local Bronze Age Andronovo culture. The Andronovo people were genetically closely related to Late Bronze Age Sintashta and Srubnaya people with significant evidence of European farmer-related ancestry [1]. Our results suggest that the Yamnaya and/or Afanasievo-related ancestry expanded further south through the Dzungarian Basin into the northern slope of the Tianshan Mountains in Xinjiang since at least the second millennium BCE and thus support the “Steppe hypothesis” for the early peopling of Xinjiang. A potential direction of future research is the generation of genome data from ancient individuals from the Tarim Basin to see whether the Steppe-related ancestry had come across the geographic barrier and spread to the south of the Tianshan mountains. Our analysis might thus provide direct evidence for the steppe-related eastern spread of Indo-European languages, notably the arrival of the Tocharian branch of Indo-European in the Tarim basin. **The Tocharian languages are known from manuscripts from 500 to 800 CE, found south of the Tianshan Mountains, in the north central and eastern reaches of the Tarim Basin. The relationship between these languages is commonly explained by divergence from a common ancestor, Proto-Tocharian, which must have preceded the attested languages by more than a millennium, probably dating to the middle 1<sup>st</sup> millennium BCE [30].** Damgaard et al. [31] pointed out that proto-Tocharian is not more closely affiliated with Indo-Iranian than with any other branch of Indo-European and that it therefore cannot be related to the Indo-Iranian Sintashta and Andronovo cultures. As the homeland of Proto-Tocharian seems to be situated south of the Tianshan Mountains, future research should determine to what extent the Tianshan has served as a geographic barrier for the ancient Steppe gene flow.

Combining both the genetic and archaeological evidence, we here provide the first direct evidence of an early stage of population admixture around 2,100 BP in Xinjiang in Western China. Our study supports the “Steppe hypothesis” over the “Bactrian Oasis hypothesis” for the peopling of the Xinjiang

region. The high amount of Yamnaya or Afanasievo-related ancestry in the Iron Age Xinjiang individuals indirectly supports the introduction of Indo-European languages into the region that survived in the form of Tocharian until the late first millennium CE. We note that we need more individuals from different sites and time periods to shed more light on the genetic history of the Tarim basin and the whole Xinjiang region.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Archaeological Context and Skeletal Materials of Shirenzigou Site
- **METHOD DETAILS**
  - Ancient DNA processing and quality control
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Genomic data processing
  - Authentication and quality control
  - Population genetic analysis
  - Testing for streams of ancestry and inference of mixture proportions
  - Genetic relatedness estimation
  - Biological sex determination
  - Y chromosomal and mitochondrial Haplogroups analysis
- **DATA AND CODE AVAILABILITY**

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2019.06.044>.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (grant no. 41472024), major project of Humanities and Social Sciences Key Research Base of the Ministry of Education (grant no. 16JJD780005), and major projects of the National Social Science Fund of China (grant no. 14ZDB052); C.-C.W. was funded by Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302), the Fundamental Research Funds for the Central Universities (ZK1144), and National Natural Science Foundation of China (31801040). C.N. and M.R. were supported by the European Union’s Horizon 2020 Research and Innovation Programme-Millet and beans, language and genes: The origin and dispersal of the Transeurasian family (granted to M.R., grant no. 646612) and the Max Planck Society.

## AUTHOR CONTRIBUTIONS

Y.C. and J.K. conceived and designed the study. J.M., Z.N., and Y.T. collected the samples and offered the <sup>14</sup>C data. S.G., Y.Y., X.Z., F.Z., and X.W. performed laboratory work. C.N. and C.-C.W. performed bioinformatic and population genetic analysis. C.N., C.-C.W., M.R., J.K., and Y.C. wrote the manuscript with input from all authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 3, 2019  
Revised: April 8, 2019  
Accepted: June 13, 2019  
Published: July 25, 2019

## REFERENCES

- Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172.
- Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374.
- Vagheesh, M., Narasimhan, N.P., Priya, M., Iosif, L., Mark, L., Swapan, M., Nadin, R., Rebecca, B., Alexander, M.K., Nathan, N., et al. (2018). The genomic formation of South and Central Asia. *bioRxiv*. <https://doi.org/10.1101/292581>.
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211.
- Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C., Ixan, O., Khussainova, E., et al. (2018). Characterizing the genetic history of admixture across inner Eurasia. *bioRxiv*. <https://doi.org/10.1101/327122>.
- Feng, Q., Lu, Y., Ni, X., Yuan, K., Yang, Y., Yang, X., Liu, C., Lou, H., Ning, Z., Wang, Y., et al. (2017). Genetic history of Xinjiang's Uyghurs suggests Bronze Age multiple-way contacts in Eurasia. *Mol. Biol. Evol.* 34, 2572–2582.
- Li, C., Ning, C., Hagelberg, E., Li, H., Zhao, Y., Li, W., Abuduresule, I., Zhu, H., and Zhou, H. (2015). Analysis of ancient human mitochondrial DNA from the Xiaohu cemetery: insights into prehistoric population movements in the Tarim Basin, China. *BMC Genet.* 16, 78.
- Gao, S., Cui, Y., Yang, Y., Duan, R., Abuduresule, I., Mair, V.H., Zhu, H., and Zhou, H. (2008). Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang, China. *Sci. China C Life Sci.* 51, 205–213.
- Zhang, F., Xu, Z., Tan, J., Sun, Y., Xu, B., Li, S., Zhao, X., Zhou, H., Gong, G., Zhang, J., and Jin, L. (2010). Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am. J. Phys. Anthropol.* 142, 314–320.
- Yao, Y.G., Lü, X.M., Luo, H.R., Li, W.H., and Zhang, Y.P. (2000). Gene admixture in the silk road region of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes Genet. Syst.* 75, 173–178.
- Hemphill, B.E., and Mallory, J.P. (2004). Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am. J. Phys. Anthropol.* 124, 199–222.
- Chen, K.T., and Hiebert, F.T. (1995). The late prehistory of Xinjiang in relation to its neighbors. *J. World Prehist.* 9, 243–300.
- Han, K. (2009). *Silk Road of Ancient Ethnic Studies*, Chinese Edition (Xinjiang People's Publishing House).
- 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.
- Wells, R.S., Yuldashesheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., et al. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98, 10244–10249.
- Zhong, H., Shi, H., Qi, X.B., Duan, Z.Y., Tan, P.P., Jin, L., Su, B., and Ma, R.Z. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* 28, 717–727.
- Wen, R., Zhao, Z.Q., Ma, J., and Wang, J.X. (2016). [Chemical analysis of ancient glass beads from Shirengigou sites in Balkun County, Xinjiang]. *Guangpuxue Yu Guangpu Fenxi* 36, 2961–2965.
- Ma, J. (2014). The rise of nomads on the Barkol Steppe and their cultural interflow with the Altai Region during the 1st millennium BCE. In *Eurasian Studies II*, Y.U. Taishan, and L.I. Jinxiu, eds. (Asia Publishing Nexus), pp. 29–37.
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503.
- Shriner, D. (2018). Re-analysis of whole genome sequence data from 279 Ancient Eurasians reveals substantial ancestral heterogeneity. *Front. Genet.* 9, 268.
- Krzewinska, M., Kilinc, G.M., Juras, A., Koptekin, D., Chylenski, M., Nikitin, A.G., Shcherbakov, N., Shuteleva, I., Leonova, T., Kraeva, L., et al. (2018). Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci. Adv.* 4, eaat4457.
- de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360. Published online June 29, 2018. <https://doi.org/10.1126/science.aar7711>.
- Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
- Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.
- Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* 536, 419–424.
- Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. *Nature* 488, 370–374.
- Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193, 1233–1254.
- Xu, S., Huang, W., Qian, J., and Jin, L. (2008). Analysis of genomic admixture in Uyghur and its implication in mapping strategy. *Am. J. Hum. Genet.* 82, 883–894.
- Carling, G. (2005). Proto-Tocharian, Common Tocharian, and Tocharian: On the value of linguistic connections in a reconstructed language. In *Proceedings of the Sixteenth Annual UCLA Indo-European Conference*, K.M. Jones-Bley, E.H.A.D. Volpe, and M.R. Dexter, eds. (Institute for the Study of Man), pp. 47–71.
- Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Rebaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppe. *Nature* 557 (Suppl 2), 10–14.
- Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88.
- Siska, V., Jones, E.R., Jeon, S., Bhak, Y., Kim, H.M., Cho, Y.S., Kim, H., Lee, K., Veselovskaya, E., Balueva, T., et al. (2017). Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* 3, e1601877.
- Ginolhac, A., Rasmussen, M., Gilbert, M.T., Willerslev, E., and Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27, 2153–2155.

35. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* *16*, 224.
36. Wang, J. (2008). The main archaeological findings from Dongheigou (Shirenzigou) site, Balikun county, Xinjiang Uygur Autonomous Region. *J. Northwe. Univ.* *38*, 36–91.
37. Tian, D., Ma, J., Wang, J., Pilgram, T., Zhao, Z., and Liu, X. (2018). Cultivation of naked barley by Early Iron Age Agro- pastoralists in Xinjiang, China. *Environ. Archaeol.* *23*, 416–425.
38. Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczyk, P., et al. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* *10*, e0129102.
39. Sirak, K.A., Fernandes, D.M., Cheronet, O., Novak, M., Gamarra, B., Balassa, T., Bernert, Z., Cséki, A., Dani, J., Gallina, J.Z., et al. (2017). A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques* *62*, 283–289.
40. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* *110*, 15758–15763.
41. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biol.* *17*, 60.
42. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
43. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* *15*, 356.
44. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* *505*, 87–91.
45. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. *Nature* *534*, 200–205.
46. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* *513*, 409–413.
47. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* *6*, 8912.
48. Jeong, C., Ozga, A.T., Witonsky, D.B., Malmström, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S., et al. (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc. Natl. Acad. Sci. USA* *113*, 7485–7490.
49. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* *514*, 445–449.
50. Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M., et al. (2017). Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am. J. Hum. Genet.* *101*, 274–282.
51. Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and Mycenaeans. *Nature* *548*, 214–218.
52. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. *Science* *345*, 1255832.
53. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* *463*, 757–762.
54. Rasmussen, M., Sikora, M., Albrechtsen, A., Korneliussen, T.S., Moreno-Mayar, J.V., Poznik, G.D., Zollikofer, C.P.E., de León, M.P., Allentoft, M.E., Moltke, I., et al. (2015). The ancestry and affiliations of Kennewick Man. *Nature* *523*, 455–458.
55. Saag, L., Varul, L., Scheib, C.L., Stenderup, J., Allentoft, M.E., Saag, L., Pagani, L., Reidla, M., Tambets, K., Metspalu, E., et al. (2017). Extensive farming in Estonia started through a sex-biased migration from the steppe. *Curr. Biol.* *27*, 2185–2193.
56. Unterländer, M., Palstra, F., Lazaridis, I., Piliipenko, A., Hofmanová, Z., Groß, M., Sell, C., Blöcher, J., Kirsanow, K., Rohland, N., et al. (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat. Commun.* *8*, 14615.
57. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J., and Fu, Q. (2017). 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr. Biol.* *27*, 3202–3208.e9.
58. Kılınç, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., et al. (2016). The demographic development of the first farmers in Anatolia. *Curr. Biol.* *26*, 2659–2666.
59. Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature* *538*, 510–513.
60. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* *538*, 201–206.
61. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* *81*, 559–575.
62. Kennet, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., et al. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* *8*, 14115.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological Samples</b>		
Shirenzigou samples	Northwest University, China	See <a href="#">Table S1</a> and <a href="#">Data S1</a> for more details
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
USER enzyme	NEB	M5505
T4 Polynucleotide Kinase	NEB	M0236L
T4 DNA Polymerase	NEB	M0203L
Bst DNA Polymerase	NEB	M0275S
Q5 High-Fidelity DNA Polymerase	NEB	M0491S
0.5M EDTA PH 8.0	Appllichem	A4892
20% SDS Solution	Serva	39575.01
1x Tris-EDTA pH 8.0	AppliChem	A8569,0500
BSA 20mg/mL	NEB	B9000 S
5 M NaCl	Sigma	S5150
Guanidine hydrochloride	Sigma Aldrich	50933
Proteinase K	Sigma Aldrich	P6556
Isopropanol	Sigma Aldrich	67-63-0
Sodium acetate	Sigma Aldrich	S2889
1M Tris-HCL, PH 8.0	Sigma Aldrich	AM9856
Tween-20	Sigma Aldrich	P9416
1M NaOH	Sigma Aldrich	71463
dNTP Mix	Thermo	R1121
ATP	Thermo	R0441
T4 DNA Ligase	Thermo	EL0011
Min Elute PCR Purification Kit	QIAGEN	28006
Agencourt AMPure XP beads (60ml)	Beckman Coulter	A63881
<b>Deposited Data</b>		
Raw and analyzed data	This paper	ENA: ERP115006
<b>Software and Algorithms</b>		
EAGER	[32]	<a href="https://github.com/apeltzer/EAGER-GUI">https://github.com/apeltzer/EAGER-GUI</a>
ANGSD	[33]	<a href="http://popgen.dk/angsd/index.php/ANGSD">http://popgen.dk/angsd/index.php/ANGSD</a>
EIGENSOFT v.6.0.1	[23]	<a href="https://github.com/DReichLab/EIG">https://github.com/DReichLab/EIG</a>
ADMIXTURE	[24]	<a href="http://software.genetics.ucla.edu/admixture">http://software.genetics.ucla.edu/admixture</a>
Samtools-0.1.19	[34]	<a href="https://sourceforge.net/projects/samtools/">https://sourceforge.net/projects/samtools/</a>
mapDamage v2.0.6	[35]	<a href="https://ginolhac.github.io/mapDamage/">https://ginolhac.github.io/mapDamage/</a>
ADMIXTOOLS	[25]	<a href="https://github.com/DReichLab/AdmixTools">https://github.com/DReichLab/AdmixTools</a>

### LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yinqiu Cui ([cuiyq@jlu.edu.cn](mailto:cuiyq@jlu.edu.cn)).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

Xinjiang is located in Central Asia has been an important corridor connecting east and west Eurasians. The findings of the Tocharian manuscripts- a branch of Indo-European language family and the well-preserved “Caucasus” feature mummies in this region thus raise questions on how and where the populations in Xinjiang come from? Two competing hypotheses has been put forward- the

“Bactrian Oasis hypothesis” and the “Steppe hypothesis,” to explain the first appearance of agricultural communities in the region ca. 2,000 BCE [11]. Until now, no consensus has been made on the origins of populations in Xinjiang.

By sequencing the first genomic data of the Shirenzigou individuals from Xinjiang, we aimed to provide a high resolution genetic evidences for the origins of prehistoric populations in the region. We hypothesize that (i) the Shirenzigou individuals were genetically admixed of the east and west Eurasians with the Yamnaya and/or Afanasievo related ancestry as the western component and the northeast Asians serves as the eastern ancestry, (ii) the Yamnaya-like ancestry for Iron Age Shirenzigou individuals and the Tocharian manuscript found from the same region provide strong evidence for the dispersal of Indo-European languages into Xinjiang.

### Archaeological Context and Skeletal Materials of Shirenzigou Site

The Shirenzigou site (Dongheigou site) is located in the eastern Tianshan Mountain in Xinjiang, Northwest China (43°31′12.8″N 93°13′44.8″E) (Data S1A) with an elevation of 2,200 m above sea level [36]. The site was first excavated by a joint team consisted of the Hami Bureau of Cultural Heritage, Barkol County’s Cultural Relics Administration and the School of Cultural Heritage of Northwest University between 2005 and 2007, and subsequently by the same team in 2009. The Shirenzigou cemetery is an early Iron Age site dated around 200 BCE (Data S1A). It lies in the platform of the forest-free region of the Tianshan mountains, a region was covered with snow and ice for the most time of the year (except for June, July and August), and in this case the region was not suitable for human habitation [18]. The excavations of a large amount of horse, goat and bull bones showed that Shirenzigou people lived a more nomadic way of life and the site was believed to be a large-scale settlement used seasonally by pastoralists [37].

From the late Bronze Age onward, the pottery and the funeral rituals of Barkol Steppe in the east Tianshan mountain region showed characteristics that were more typical of the oasis-based agricultural Yanbulake Culture of the Hami Basin. The Shirenzigou site also resembles the Pazyryk Culture, an early Iron Age culture from the Altai mountain [18]. For example, horse sacrifice and animal motifs - golden plaques of tiger and rams- excavated from the Shirenzigou site were also popular among the Pazyryk people. Moreover, a bronze cauldron found in dwelling F2 may also indicate additional influences from the Han Chinese. In all, the current archaeological evidence shows that Shirenzigou people have a close relationship with populations of the Altai mountain as well as that of the Han Chinese.

Ten samples from the Shirenzigou site were reported in this study (Table 1), 9 individuals were from the burial tombs and the other one individual (F004) was from the dwelling. Five dated samples from this study (M4, M8R1, M820, M010 and M012) yielded uncalibrated dates of  $2215 \pm 25\text{BP}$ ,  $2230 \pm 25\text{BP}$ ,  $2185 \pm 30\text{BP}$ ,  $2290 \pm 25\text{BP}$  and  $2195 \pm 20\text{BP}$  (Data S1A). All of those five individuals were dated around 200 BCE, corresponding to the Warring State Period or the Western Han Dynasty of the Chinese chronology. We note that five individuals in our study are not directly dated, but we believe they are all contemporaneous to the other five dated samples for the following reasons:

1. These 10 individuals are buried with the same funeral ritual: stone kurgan on the surface, rectangular tomb pit with animal sacrifice, and the dead are buried in extended supine position in stone or wooden coffins with same type of grave goods.
2. All the graves are distributed in a line from the slope down which shows that these individuals have a close relationship.
3. Based on stratigraphy, all the 10 individuals were excavated from the same layer.
4. Some other individuals from this site that are not included in this study were also radiocarbon dated to the same time period of the five dated individuals from this study. Further a number of other materials such as animal bones, wood remains as well as barley grains were radiocarbon dated and all are within the same time period as the five dated individuals of this study [37].

## METHOD DETAILS

### Ancient DNA processing and quality control

#### Ancient DNA lab work

Six petrous bones and four teeth were collected from the Shirenzigou site. All samples were decontaminated by wiping the surfaces with 5% bleach before they were transferred to a dedicated clean room. In order to yield the highest amount of endogenous DNA we processed the petrous bones according to the procedures described in [38, 39], then followed by UV irradiation on both sides of the samples. DNA extractions were carried out in the clean-room facilities at Jilin University, using an in-solution silica-based protocol [40]. Briefly, 50mg of bone powder from the inner part of petrous bones and teeth were digested in 900ul 0.5M EDTA, 16.7ul of Proteinase K (15mg/ml) and 83.3ul ddH<sub>2</sub>O at 37°C for 18 hours. The supernatant was then transferred to a MinElute silica spin column (QIAGEN) after fully mixed with the 13ml custom binding buffer [5M Guanidine hydrochloride (MW 95.53), 40% Isopropanol, 90mM Sodium Acetate (3M) and 0.05% Tween-20] followed by two washes with PE buffer (80% ethanol). The DNA was then eluted with 100ul TET buffer (1M Tris-HCl, 0.5M EDTA and 10% Tween-20).

The double-stranded libraries of Shirenzigou individuals were prepared by using two different methods. Two libraries (M819 and M820) were built with the NEBNext Ultra DNA Library Prep Kit following the manufacturer’s instructions, but with minor corrections. Specifically, we diluted the NEBNext Adaptor for Illumina (15 μM) to 1.5 μM with a 10- fold dilution (1:9) in sterile water for immediate use. The specialty of the NEBNext Ultra DNA Library Prep Kit is that the adaptor provided by this kit contains a Uracil base that needs to be cleaved by the USER enzyme before the amplification step of the libraries. However, given that ancient DNA also contains Uracil bases especially at 5’ end, USER enzyme will also remove any Uracil residuals at 5’ end and thus resulted in the absence of

deamination damage on 5' end for the two samples. In contrast, the libraries for rest samples (F004, M4, M8R1, M010, M012, M15-1, M15-2, X3) were built follow the procedures described by Dabney et al. [40]. The libraries were then amplified either by Q5 High-Fidelity DNA Polymerase (New England Biolabs) (M819, M820) or by Bst enzyme (New England Biolabs) (F004, M4, M8R1, M010, M012, M15-1, M15-2, X3), and were then sent to Novogene Company (Beijing, China) to be sequenced on an Illumina HiSeq X10 platform (2x150 bp).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Genomic data processing

The raw fastq files were processed in EAGER v1.92.50 program, an automated computational pipeline specially designed for ancient DNA data processing [41]. Adapters were trimmed with AdapterRemoval v2.2.0 [32] and read length shorter than 30bp were discarded. The trimmed data was then mapped to the human reference genome (GRCh37) using BWA 0.7.12 with '-n 0.01' and '-l 1024' to allow for more mismatches and to disable the seeding. The duplicated reads were then removed with dedup v0.12.2 [41] and sequences with a mapping quality of  $\geq 30$  are retained using SAMtools [42]. To minimize the bias due to ancient DNA deamination, we trimmed the alignment reads according to the frequencies of C to T or G to A at both 5' and 3' ends to a degree that the damages at the end of the trimmed reads were identical to the baseline. We randomly called genotype for an SNP from trimmed reads with high-quality base ( $Q > 30$ ) that implemented using pileupCaller (<https://github.com/stschiff/sequenceTools>).

### Authentication and quality control

We used three methods to measure the authenticity of ancient DNA data. First, we applied mapDamage v2.0.6 [34], with default parameters to determine the molecular damage which is typical of ancient DNA (Data S1A). Second, we estimated the contamination of mitochondrial sequences using schmutzi [35], an iterative likelihood-based method that jointly estimation of ancient DNA contamination and endogenous mitochondrial consensus sequences. Third, we assessed the contamination of the nuclear genome in males using ANGSD v 0.910 [43] by estimating the X chromosome polymorphism rate on male samples with sufficient markers (Here we used at least 100 SNPs overlap with X chromosome as a threshold) (Data S1A).

### Population genetic analysis

The samples were first merged with published ancient genome-wide data for the 1240k panel [1, 2, 19, 22, 26, 33, 44–58]. The merged dataset was then compared to modern populations in the Affymetrix Human Origins (HO) public dataset [26, 59] or the high-coverage Simons Genome Diversity Project [5, 60] and the final dataset consists of 593,124 autosomal SNPs. The principal component analysis was carried out using the "Isqproject" options in the smartpca program [23]. We project our ancient samples onto the variation of present-day Eurasians from published Human Origin dataset over 593,124 SNPs. We carried out ADMIXTURE analysis [24] after pruning for linkage disequilibrium in PLINK [61] with parameters--indep-pairwise 200 25 0.4 which retained 304,935 SNPs.

### Testing for streams of ancestry and inference of mixture proportions

We used qpAdm as implemented in ADMIXTOOLS [25] to estimate mixture proportions for a *Test population* as a combination of *N reference* populations by exploiting (but not explicitly modeling) shared genetic drift with a set of outgroups. The *reference* or the *source* populations should not be equally related to the *outgroups*, and the *outgroups* must be related to the *test* populations via the references. This method provides a formal test for whether the *reference* populations provide a good fit ( $p > 0.5$  serves as a good fit) to the *test* population. This method is sensitive to the sample size as well as the coverage of the *test*, *source* and *outgroups*, usually a test with low coverage and small population size will result in a higher P value and a larger standard error. We modeled Shirengigou individuals using the following outgroups: Dinka, Ust\_Ishim, Kostenki14, Onge, Papuan, Australian, Iran\_N, EHG, LBK\_EN (Data S2B1–S2B7).

### Genetic relatedness estimation

We estimated the genetic relatedness between Shirengigou individuals using pairwise mismatch rate analysis ('pmr') [62]. The pmr value is calculated by the number of SNPs for which two individuals have different alleles divided by the number of all SNPs for which two have non-missing genotypes. In this sense, the pmr between identical individuals ( $r = 1$ ) is expected to be a half of that between two unrelated individuals ( $r = 0$ ) and pmr value for the first- ( $r = 0.5$ ) and second-degree relatives ( $r = 0.25$ ) are 3/4 and 7/8 of that between unrelated pairs (See also Figure S1).

### Biological sex determination

We assessed the biological sex of newly reported samples by evaluating both the ratios of reads aligned to the Y chromosome (yCov) and to X chromosome (xCov) with compared to the total number of reads aligning to the autosomes (autoCov) (Data S1A). Then the separation between the two ratios (xCov/autoCov and yCov/autoCov) was used to determine the genetic sex of our individuals (See also Figure S2).

### **Y chromosomal and mitochondrial Haplogroups analysis**

Y chromosome haplogroups were examined by aligning a set of positions in the ISOGG (<http://isogg.org/>) and Y-full (<https://www.yfull.com/tree/>) databases, in which we only restrict our analysis to reads with base and mapping quality higher than 30. The Y haplogroups were then determined with the most derived SNPs as well as upstream ancestral SNPs. We called mtDNA SNPs with the SAMtools mpileup tool against the rCRS (Revised Cambridge Reference Sequence) also with a minimum base and mapping quality of 30, and double checked the SNPs by visual inspection with SAMtools tview. Then, mtDNA sequences were assigned to different haplogroups using MitoTool (*phylotree build 16*) ([Data S1A](#)).

### **DATA AND CODE AVAILABILITY**

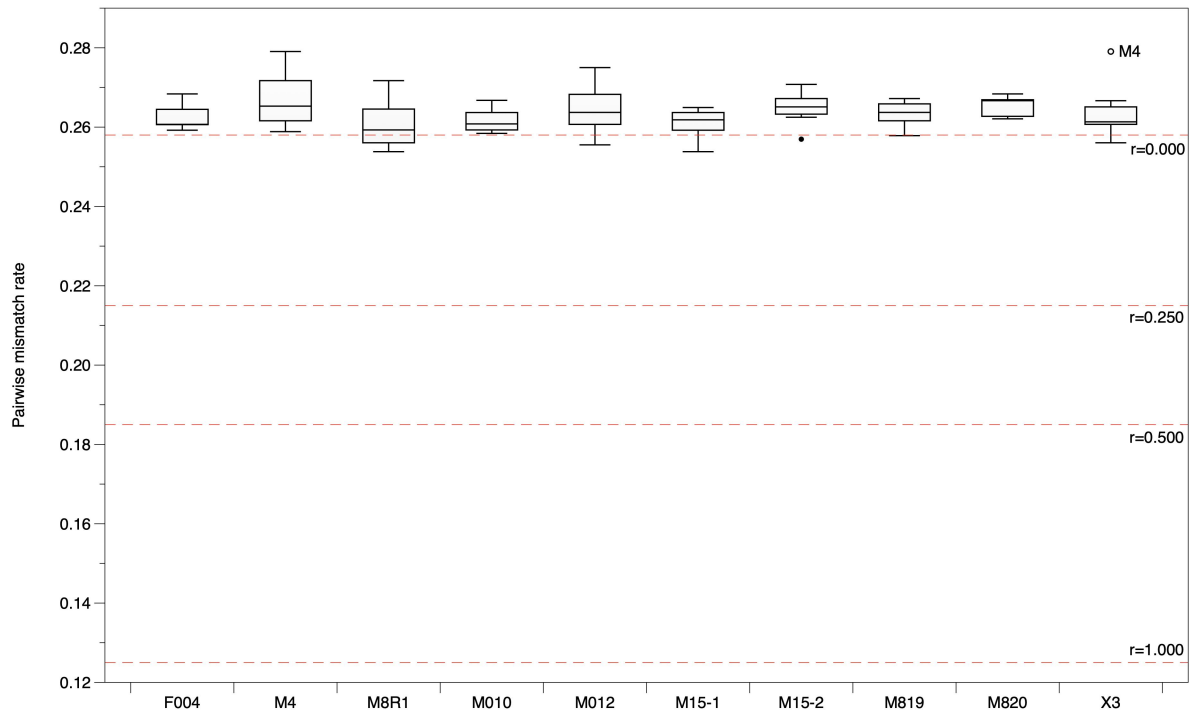
The accession number for the data reported in this paper is ENA: ERP115006.

**Current Biology, Volume 29**

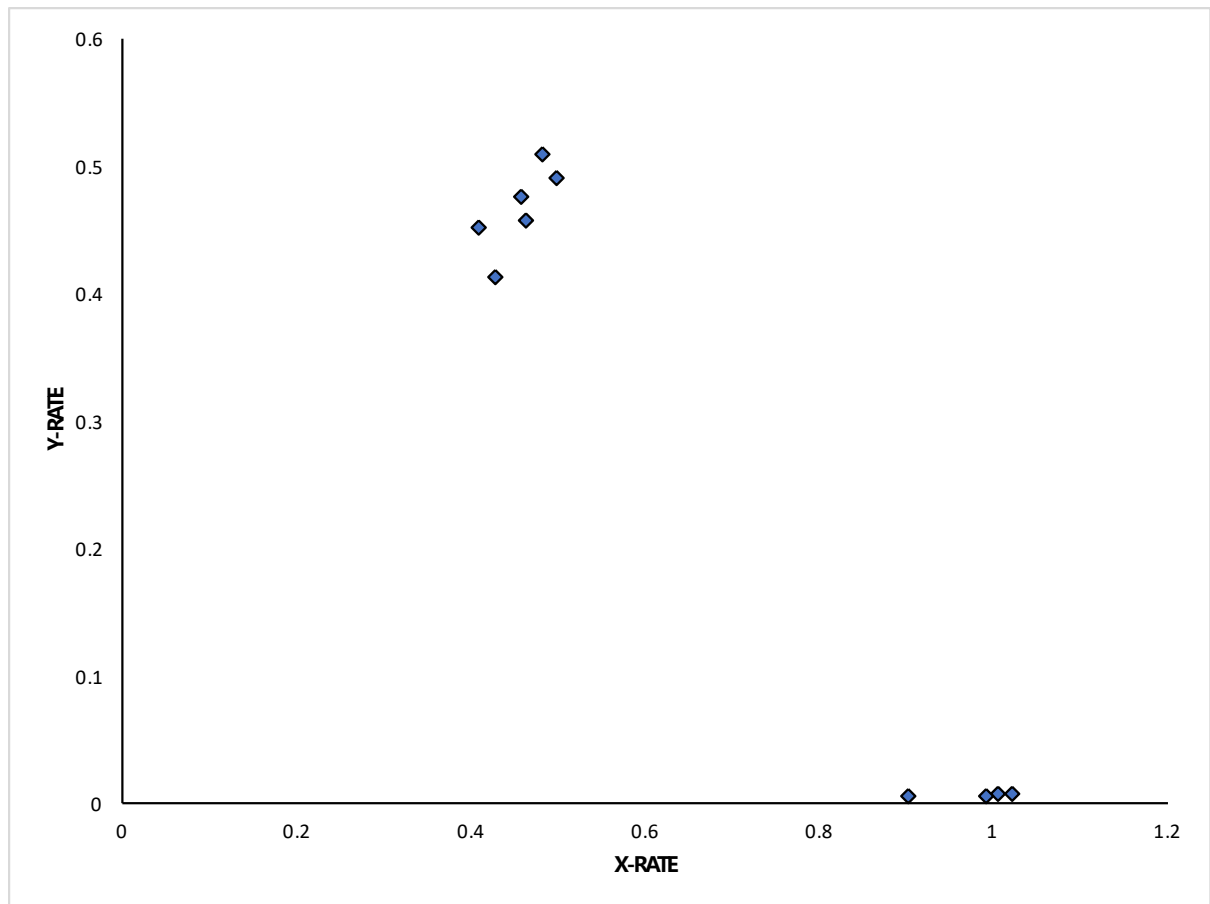
**Supplemental Information**

**Ancient Genomes Reveal Yamnaya-Related  
Ancestry and a Potential Source  
of Indo-European Speakers in Iron Age Tianshan**

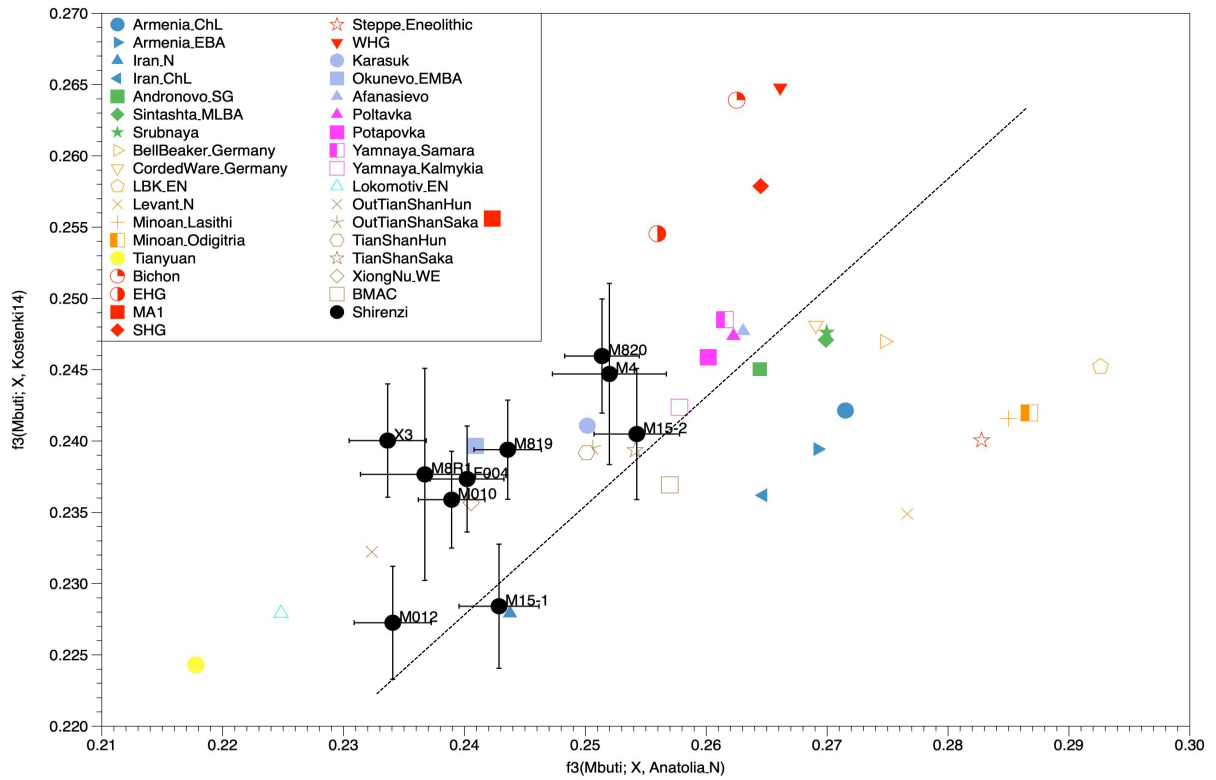
**Chao Ning, Chuan-Chao Wang, Shizhu Gao, Yang Yang, Xue Zhang, Xiyan Wu, Fan Zhang, Zhongzhi Nie, Yunpeng Tang, Martine Robbeets, Jian Ma, Johannes Krause, and Yinqiu Cui**



**Figure S1. Pairwise mismatch rate (pmr) among different Shirenzigou individuals, Related to STAR Methods.** The red dashed line marked the expected pairwise mismatch rate for the given coefficient of relationship ( $r$ ). The bottom line is the expected pmr value of the identical pairs ( $r=1$ ) and the middle red lines corresponding to the expected value for the unrelated ( $r=0$ ), first degree relative ( $r=0.5$ ) and second degree relative ( $r=0.25$ ) respectively. Each boxplot represents the variation of pmrs estimated between a selected individual (x-axis) and the rest individuals from this study. If an individual is first- or second-degree relative to the selected individual, we would expect an outlier (a dot) is on or around the dashed lines ( $r=0.5$  and  $r=0.25$ ).

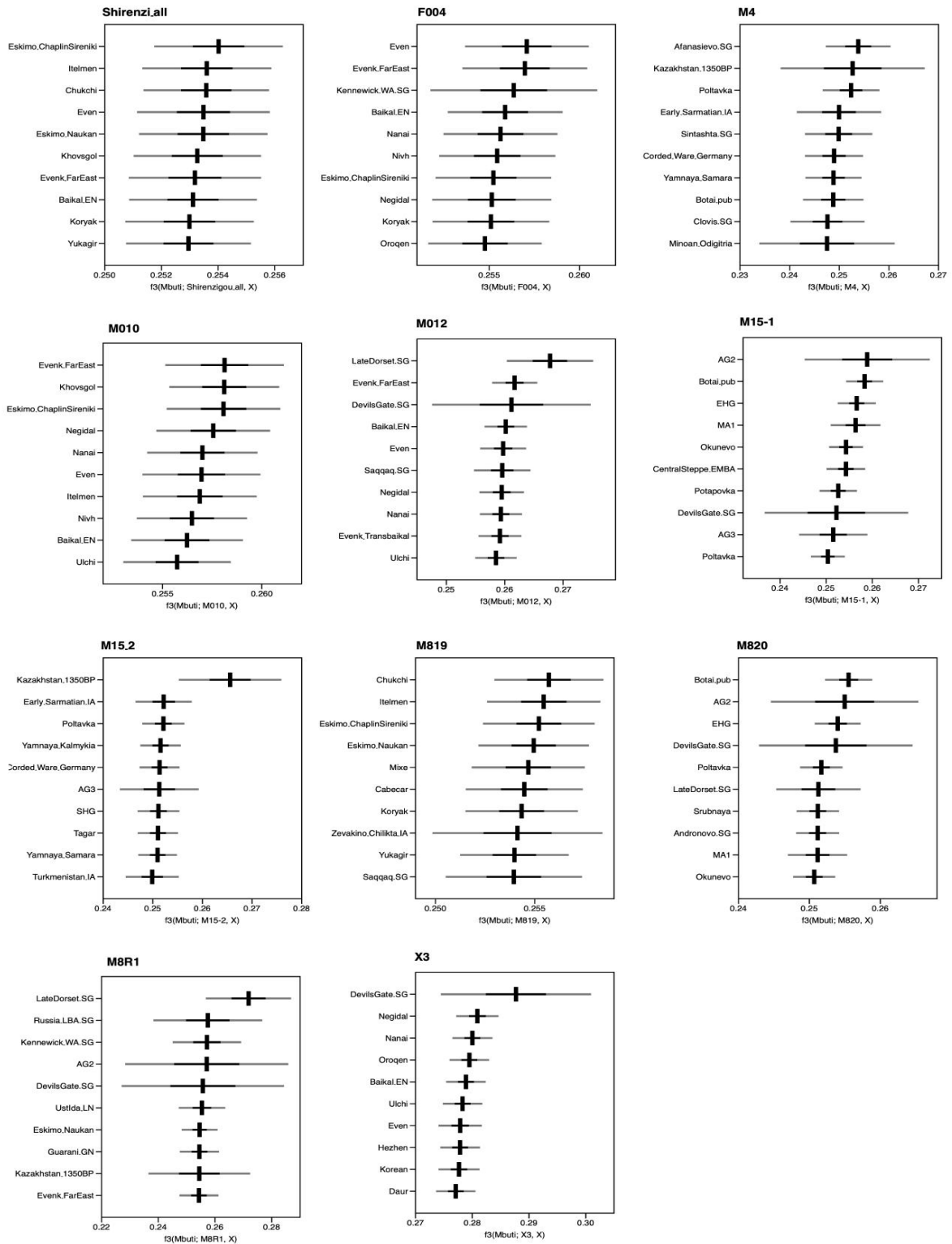


**Figure S2. Scattered plot for inferring the genetic sex of Shirenzigou individuals, Related to STAR Methods and Table 1.** “X-RATE” and “Y-RATE” are the reads mapped on X chromosome and Y chromosome with compared to that mapped on autosomes. Two clear clusters were observed which six individuals were determined as male and four are females.



**Figure S3. Biplot of  $f_3$ -outgroup tests illustrating the Kostenki14 and Anatolia\_N like ancestries in Shirenzigou individuals, Related to Figure 2.** Most Shirenzigou individuals were on a cline with Yamnaya and European hunter-gatherer groups, lacking the European farmer ancestry as compared to the Steppe\_MLBA populations such as Andronovo, Srubnaya and Sintashta [S1-S5]. Horizontal and vertical bars represent  $\pm 3$  standard errors, corresponding to form of outgroup  $f_3$  tests on the x axis and y axis respectively.





**Figure S4. Shared Genetic drift between Shirenzigou individuals and world-wide representative populations, Related to Figure 2. Shirenzl\_all represents a group that include all the Shirenzigou individuals in this study.**

Source 1	Source 2	Target	f3	std.err	Z	SNPs
Yamnaya_Samara	Korean	SRZ	-2.64E-02	1.10E-03	-24.122	217706
Poltavka	Korean	SRZ	-2.59E-02	1.09E-03	-23.817	214454
EHG	Korean	SRZ	-2.58E-02	1.31E-03	-19.812	196443
Srubnaya	Korean	SRZ	-2.56E-02	1.06E-03	-24.245	219598
Srubnaya	Negidal	SRZ	-2.53E-02	1.10E-03	-23.04	211079
Poltavka	Lokomotiv_EN	SRZ	-2.52E-02	1.39E-03	-18.214	172201
Srubnaya	Nanai	SRZ	-2.52E-02	1.03E-03	-24.347	224553
EHG	XiongNu	SRZ	-2.51E-02	1.51E-03	-16.68	170037
Corded_Ware_Germany	Korean	SRZ	-2.50E-02	1.10E-03	-22.775	214601
Corded_Ware_Germany	Lokomotiv_EN	SRZ	-2.50E-02	1.38E-03	-18.135	171574
Srubnaya	Evenk_FarEast	SRZ	-2.49E-02	1.15E-03	-21.729	206026
Poltavka	Nanai	SRZ	-2.49E-02	1.06E-03	-23.495	220061
Srubnaya	XiongNu	SRZ	-2.49E-02	1.21E-03	-20.509	202827
Poltavka	XiongNu	SRZ	-2.48E-02	1.24E-03	-20.096	195509
Yamnaya_Samara	Evenk_FarEast	SRZ	-2.48E-02	1.22E-03	-20.398	203895
Corded_Ware_Germany	Evenk_FarEast	SRZ	-2.48E-02	1.21E-03	-20.454	199415
Srubnaya	Lokomotiv_EN	SRZ	-2.48E-02	1.32E-03	-18.793	179327
Corded_Ware_Germany	Shamanka_EN	SRZ	-2.47E-02	1.08E-03	-22.81	211113
Yamnaya_Samara	Yi	SRZ	-2.47E-02	1.05E-03	-23.486	226447
Andronovo.SG	Korean	SRZ	-2.47E-02	1.14E-03	-21.738	212642
Yamnaya_Samara	Nanai	SRZ	-2.46E-02	1.07E-03	-23.033	222934
Yamnaya_Samara	Naxi	SRZ	-2.46E-02	1.06E-03	-23.204	224856
Yamnaya_Samara	Tujia	SRZ	-2.46E-02	1.05E-03	-23.373	226414
Yamnaya_Samara	She	SRZ	-2.46E-02	1.07E-03	-23.078	224556
Poltavka	Shamanka_EN	SRZ	-2.45E-02	1.07E-03	-22.861	210900
Yamnaya_Samara	Han	SRZ	-2.45E-02	1.04E-03	-23.585	235957
Corded_Ware_Germany	Nanai	SRZ	-2.44E-02	1.09E-03	-22.385	220557
Yamnaya_Samara	Korean	SRZ	-2.64E-02	1.10E-03	-24.122	217706
Poltavka	Korean	SRZ	-2.59E-02	1.09E-03	-23.817	214454
EHG	Korean	SRZ	-2.58E-02	1.31E-03	-19.812	196443
Srubnaya	Korean	SRZ	-2.56E-02	1.06E-03	-24.245	219598
Srubnaya	Negidal	SRZ	-2.53E-02	1.10E-03	-23.04	211079
Poltavka	Lokomotiv_EN	SRZ	-2.52E-02	1.39E-03	-18.214	172201
Srubnaya	Nanai	SRZ	-2.52E-02	1.03E-03	-24.347	224553
EHG	XiongNu	SRZ	-2.51E-02	1.51E-03	-16.68	170037
Corded_Ware_Germany	Korean	SRZ	-2.50E-02	1.10E-03	-22.775	214601
Corded_Ware_Germany	Lokomotiv_EN	SRZ	-2.50E-02	1.38E-03	-18.135	171574
Srubnaya	Evenk_FarEast	SRZ	-2.49E-02	1.15E-03	-21.729	206026
Poltavka	Nanaia	SRZ	-2.49E-02	1.06E-03	-23.495	220061
Srubnaya	XiongNu	SRZ	-2.49E-02	1.21E-03	-20.509	202827

**Table S1.** The top 40 most negative Z scores of admixture f3 statistics in the form of (*Source 1*, *Source 2* ; *Target*) using Human Origins dataset, Related to Figure 2. SRZ represents all the Shirenzigou individuals in this study.

Sample	Bustan_BA	Sappali_Tepe_BA	Andronovo.SG	Srubnaya
F004	3.581E-08	1.051E-11	9.919E-03	1.841E-04
M010	5.872E-12	5.978E-14	1.261E-03	1.158E-05
M012	6.511E-03	1.792E-04	3.252E-02	5.544E-03
M15-1	2.353E-19	3.563E-21	2.030E-03	2.942E-06
M15-2	3.380E-09	1.498E-12	5.877E-01	1.013E-02
M4	1.146E-03	7.972E-05	3.380E-02	2.806E-04
M819	4.928E-29	1.322E-36	1.698E-05	2.599E-07
M820	2.058E-16	5.494E-25	7.596E-04	2.604E-06
M8R1	9.411E-03	1.856E-03	2.460E-01	1.002E-01
X3	5.407E-04	5.471E-05	6.578E-02	2.574E-02

**Table S2. P values in modelling a two-way (P=rank 1) admixture in Shirenzigou samples using each of the four populations (*Bustan\_BA*, *Sappali\_Tepe\_BA*, *Andronovo.SG*, *Srubnaya*) together with Han Chinese as two sources [S6], Related to Figure 2. We used the following set of outgroups populations: Dinka, Ust\_Ishim, Kostenki14, Onge, Papuan, Australian, Iran\_N, EHG, LBK\_EN.**

Pop1	Pop2	Pop3	Pop4	F4	Z	nBABA	nABBA	nSNPs
Han	Itelmen	SRZ	Mbuti	-1.05E-03	-5.545	24192	24736	516512
Han	Even	SRZ	Mbuti	-1.03E-03	-4.99	23954	24485	516505
Han	Shamanka_EN	SRZ	Mbuti	-9.40E-04	-6.077	24065	24551	516501
Han	Evenk_FarEast	SRZ	Mbuti	-9.29E-04	-4.908	23599	24072	509666
Han	Koryak	SRZ	Mbuti	-9.12E-04	-5.062	24215	24686	516512
Han	Nganasan	SRZ	Mbuti	-8.81E-04	-5.518	24132	24587	516512
Han	Negidal	SRZ	Mbuti	-8.75E-04	-5.025	23585	24031	509670
Han	Nanai	SRZ	Mbuti	-8.48E-04	-6.686	23599	24031	509670
Han	Evenk_Transbaikal	SRZ	Mbuti	-7.97E-04	-5.284	24047	24453	509670
Han	Shamanka_EBA	SRZ	Mbuti	-7.18E-04	-3.654	24233	24602	513727
Han	Oroqen	SRZ	Mbuti	-7.13E-04	-5.799	23723	24091	516512
Han	Yakut	SRZ	Mbuti	-6.79E-04	-5.36	24211	24562	516512
Han	Ulchi	SRZ	Mbuti	-6.45E-04	-5.718	23782	24116	516512
Han	Nivh	SRZ	Mbuti	-5.79E-04	-3.991	23726	24021	509670
Han	Hezhen	SRZ	Mbuti	-4.64E-04	-3.831	23715	23955	516512
Han	Korean	SRZ	Mbuti	-3.89E-04	-3.214	23462	23663	516511
Han	Buryat	SRZ	Mbuti	-2.99E-04	-2.897	24270	24422	509670
Han	Xibo	SRZ	Mbuti	-2.61E-04	-2.477	23722	23857	516512
Han	Japanese	SRZ	Mbuti	-2.01E-04	-2.586	23583	23687	516512
Han	Yi	SRZ	Mbuti	-7.00E-05	-0.715	23706	23742	516512
Han	She	SRZ	Mbuti	-1.10E-05	-0.119	23473	23479	516512
Han	Tujia	SRZ	Mbuti	1.70E-05	0.185	23523	23514	516512
Han	Naxi	SRZ	Mbuti	6.40E-05	0.602	23741	23708	516512
Han	Tu	SRZ	Mbuti	9.80E-05	0.966	23950	23900	516512
Han	Tibetan	SRZ	Mbuti	3.21E-04	1.578	23878	23715	509681
Han	Kinh	SRZ	Mbuti	5.11E-04	4.585	23757	23493	516512
Han	Ami	SRZ	Mbuti	5.56E-04	4.394	23655	23368	516512
Han	Dai	SRZ	Mbuti	6.06E-04	5.938	23727	23414	516512
Han	Atayal	SRZ	Mbuti	7.86E-04	5.42	23757	23351	516512
Han	Uygur	SRZ	Mbuti	1.43E-03	9.432	25841	25105	516512
Han	Thai	SRZ	Mbuti	1.43E-03	12.597	24540	23802	516512
Han	Cambodian	SRZ	Mbuti	1.60E-03	12.936	24550	23726	516512
Han	Kumyk	SRZ	Mbuti	3.37E-03	13.761	27947	26204	516512
Han	Onge	SRZ	Mbuti	5.02E-03	20.455	27002	24410	516512
Han	Papuan	SRZ	Mbuti	7.08E-03	23.801	28017	24361	516512

**Table S3. Symmetry test between Shirezigou people and Han Chinese with respect to other representative east Eurasians, Related to Figure 2.** Shirezigou people shares excess alleles with populations Northeast Eurasians. Note: SRZ, Shirezigou individuals in this study.

## Supplemental References

- S1. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* *522*, 207-211.
- S2. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* *505*, 87-91.
- S3. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* *536*, 419-424.
- S4. Fu, Q.M., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwangler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. *Nature* *534*, 200-205.
- S5. Allentoft, M.E., Sikora, M., Sjogren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlstrom, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* *522*, 167-172.
- S6. Vagheesh M., Narasimhan, N.P., Priya M., Iosif L., Mark L., Swapan M., Nadin R., Rebecca B., Alexander M.K., Nathan N., et al. (2018). The Genomic Formation of South and Central Asia. *bioRxiv*. <http://dx.doi.org/10.1101/292581>.