


Genetic Continuity of Bronze Age Ancestry with Increased Steppe-Related Ancestry in Late Iron Age Uzbekistan

Vikas Kumar,^{1,2} E. Andrew Bennett,¹ Dongyue Zhao,³ Yun Liang,³ Yunpeng Tang,³ Meng Ren,³ Qinyan Dai,¹ Xiaotian Feng,¹ Peng Cao,¹ Ruowei Yang,¹ Feng Liu,¹ Wanjing Ping,¹ Ming Zhang,^{1,4} Manyu Ding,^{1,4} Melinda A. Yang,⁵ Berdimurodov Amridin,⁶ Hasanov Mutalib,^{*,6} Jianxin Wang,^{*,3} and Qiamei Fu ^{*,1,2,4}

¹Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, China

²Shanghai Qi Zhi Institute, Shanghai 200232, China

³School of Cultural Heritage, Northwest University, Xi'an, China

⁴University of Chinese Academy of Sciences, Beijing, China

⁵Department of Biology, University of Richmond, Richmond, VA, USA

⁶Institute of Archaeology, Academy of Sciences of Uzbekistan, Samarkand, Uzbekistan

*Corresponding authors: E-mails: khasanov.mutalib@mail.ru; moguawang@nwu.edu.cn; fuqiamei@ivpp.ac.cn.

Associate editor: Maria C. Ávila-Arcos

Abstract

Although Uzbekistan and Central Asia are known for the well-studied Bronze Age civilization of the Bactria–Margiana Archaeological Complex (BMAC), the lesser-known Iron Age was also a dynamic period that resulted in increased interaction and admixture among different cultures from this region. To broaden our understanding of events that impacted the demography and population structure of this region, we generated 27 genome-wide single-nucleotide polymorphism capture data sets of Late Iron Age individuals around the Historical Kushan time period (~2100–1500 BP) from three sites in South Uzbekistan. Overall, Bronze Age ancestry persists into the Iron Age in Uzbekistan, with no major replacements of populations with Steppe-related ancestry. However, these individuals suggest diverse ancestries related to Iranian farmers, Anatolian farmers, and Steppe herders, with a small amount of West European Hunter Gatherer, East Asian, and South Asian Hunter Gatherer ancestry as well. Genetic affinity toward the Late Bronze Age Steppe herders and a higher Steppe-related ancestry than that found in BMAC populations suggest an increased mobility and interaction of individuals from the Northern Steppe in a Southward direction. In addition, a decrease of Iranian and an increase of Anatolian farmer-like ancestry in Uzbekistan Iron Age individuals were observed compared with the BMAC populations from Uzbekistan. Thus, despite continuity from the Bronze Age, increased admixture played a major role in the shift from the Bronze to the Iron Age in southern Uzbekistan. This mixed ancestry is also observed in other parts of the Steppe and Central Asia, suggesting more widespread admixture among local populations.

Key words: ancient DNA, population genomics, evolution.

Introduction

Uzbekistan, in Central Asia, includes diverse populations with both East and West Eurasian ancestries (Irwin et al. 2010). During the Bronze Age (BA), the emergence of pastoral economies led to an increased mobility and the development of many settlements across Central Asia (Kohl 2007; Frachetti 2012; Spengler 2015). Many such pastoral settlements from the Mid-BA (~2100–1800 BCE), particularly in and around South Uzbekistan, possess a material culture associated with the Bactria–Margiana Archaeological Complex (BMAC) (Kohl 2007; Cunliffe 2015). In particular, BA Uzbekistan populations show similar ancestry profiles to those found in other BMAC settlements around the Amu Darya River, suggesting interactions and

connections between them (Narasimhan et al. 2019). The BMAC populations were previously shown to be primarily a mixture of Iranian (~60–65%) and Anatolian (~20–25%) farmer ancestries (Narasimhan et al. 2019). Some BMAC individuals were found to have high Yamnaya/Steppe-related ancestry, suggesting this ancestry began appearing in Central Asia by around ~4100 BP (Narasimhan et al. 2019). Later, in the Mid and Late BA, communities residing in the Bactrian region of Uzbekistan showed higher Steppe-related ancestry compared with the Early BA populations, suggesting an increased influence from the Steppe herders in Uzbekistan (Narasimhan et al. 2019).

In Central Asia, the transition from the BA to the Iron Age (IA) toward the middle of the second millennium BCE is

characterized by major shifts in material culture with increased mobility and interaction (Cunliffe 2015). During the Early IA, South-Central Asia contained many important settlements, such as Bactria, located in the Amu-Darya basin, Margiana, located near the delta of the Murghab River, Sgodia, around the Zeravshan River basin, and Fergana valley, near the Tianshan mountain range (Lhuillier and Mashkour 2017). During the beginning of the IA (~1500–1400 BCE) a new culture, “Yaz,” appeared in South-Central Asia characterized by a shift in material culture that included a distinct hand-made pottery and different funerary practices, which lacked prominent graves like the Kurgan-style burial graves observed in the Steppe region (Sarmiento and Lhuillier 2011; Lhuillier and Mashkour 2017). Later, during the middle IA, which extended from ~1000 BCE until the conquest of Central Asia by the Achaemenids, the Persian Empire saw the development of large settlements, advancement in Iron metallurgy technology and with a new wheel-made pottery style by the “Yaz II” culture, followed by the late IA establishment of the Achaemenid Empire (~550–330 BCE), also sometimes referred to as “Yaz III” (Kuhrt 2001; Wu et al. 2015; Lhuillier and Mashkour 2017).

Later, this region was under the control of the Greco-Bactrian-Kingdom (~250–125 BCE) (Abazov 2008), during which major upheavals occurred, with migrations and settlements by Indo-European tribes such as the Sakas from North Asia and later by the Yuezhi people from East Asia, who went on to establish the Kushan Empire (~1st century CE) (Abdullaev 2007; Abazov 2008; Cunliffe 2015; Lhuillier and Mashkour 2017). In general, in archaeological and genetic studies, the impact of Steppe-related culture and ancestry increased from the BA to IA in Central Asia (Cunliffe 2015; Narasimhan et al. 2019), thus genetically, a question remains of how Steppe-related migration into this region from the BA impacted the late IA populations in Central Asia. To address this question, and the general lack of post-BA genomic data from Uzbekistan, we sequenced 27 ancient samples from the Late IA encompassing the Kushan time period (Uz_IA) (~2100–1500 BP), from three sites in South Uzbekistan: Rabat, Serkharakat, and Dehkan (fig. 1A and supplementary table S1, Supplementary Material online).

Results and Discussion

In total, DNA from 27 ancient individuals was extracted and enriched for ~1.2 million single-nucleotide polymorphism (SNP) targets (Fu et al. 2013; 2015). Full mitochondrial genomes were additionally enriched from all 27 samples using oligonucleotide probes (Fu et al. 2013) (supplementary table S1, Supplementary Material online). Mitogenomes from all 27 individuals (19–370 × fold coverage) had low contamination rates (0–2.7%) and a high haplogroup diversity among the Uz_IA population (fig. 1A and supplementary table S1 and fig. S15, Supplementary Material online). The haplogroups suggest ancient Turan (i.e., the region comprising Iran and southern Central Asia, which includes present-day countries of Turkmenistan, Tajikistan, Uzbekistan, Afghanistan, and Kyrgyzstan) and Steppe-related maternal ancestry

represented by six “U” and five “H” haplogroups (Narasimhan et al. 2019). Both of these haplogroups are also prevalent in present-day European, Caucasus, and Central and West Asian populations (supplementary fig. S15, Supplementary Material online). Another four D-haplogroups indicate a stronger genetic connection with present-day East Asian and Siberian populations (supplementary fig. S15, Supplementary Material online). Twelve other minor haplogroups of J, K, M, T, V, W, and X also indicate a high haplogroup diversity widely present in this region both in the ancient and present-day populations (Irwin et al. 2010; Narasimhan et al. 2019) (fig. 1A and supplementary table S1, Supplementary Material online). We also identified the Y-haplogroup, R1a1 ($n = 3$), which additionally supports a connection to the present-day West-Central Asian and eastern European populations. Among the ancient populations, this haplogroup has been reported in Middle-to-Late Bronze Age (MLBA) populations in cultures having Steppe-related affinities such as Corded-Ware, Andronovo, and Sintashta (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015; Underhill et al. 2015). We used nuclear DNA to further investigate the genetic relationships of IA Uzbekistan. We discarded individuals with <20k SNPs and checked the kinship relationships of those remaining, using only the individual with the highest number of SNPs from each kinship pair for further genetic analysis (supplementary table S2, Supplementary Material online). This resulted in a set of 15 unrelated individuals having 0.02–2.88 × fold coverage (27,900–723,918 SNPs).

Increased Steppe Ancestry in Iron Age Compared with Bronze Age

To assess the genetic affinities between Uz_IA individuals and ancient populations, we first performed principal component analysis (PCA) (Patterson et al. 2006) by projecting the ancient diverse Eurasian populations onto the present-day populations. The PCA results show that all the Uz_IA individuals lie on a cline extending from West Siberian Hunter Gatherers (WSHG) to Anatolian and Iranian farmer-related ancestry (fig. 1B and supplementary figs. S1–S3, Supplementary Material online) clustering closely with the geographically proximal ancient BA and IA populations from the Steppe and Central Asia, including ancient populations from Uzbekistan and neighboring regions of Turan and the Central Steppe (fig. 1B and supplementary figs. S1 and S2, Supplementary Material online). Despite being in proximity on the PCA to the previously published Uzbekistan BA populations (Bustan, Dzarkutan, Kashkarchi, Kokcha, and Sappali_Teppe), the Uz_IA populations lack a tight clustering. They overlap with Turan IA and historical time period populations with Steppe-related ancestry, suggesting the presence and influence of more Steppe-related ancestry in Uz_IA individuals than in BA Uzbekistan populations. To estimate the ancestral admixture and population structure, we used the model-based clustering method implemented in ADMIXTURE (Alexander et al. 2009). The same trends from the PCA are also observed in the ADMIXTURE ($K = 8$) analysis (fig. 1C and supplementary figs. S4 and S5, Supplementary

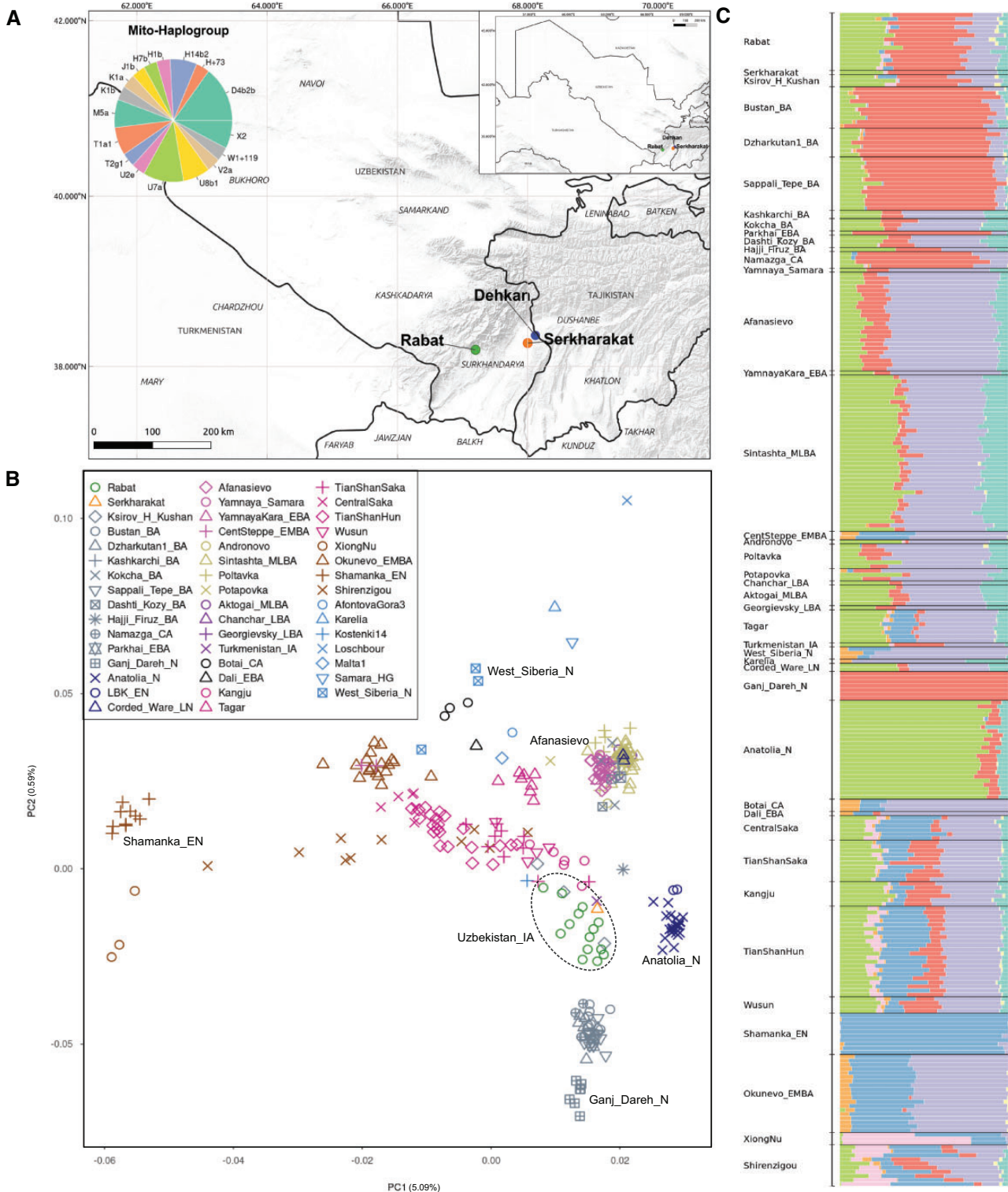


FIG. 1. Geographical location, mitochondrial haplogroup, PCA and admixture analysis of Uz_IA individuals. (A) Map showing the geographical location in South Uzbekistan of the three Uz_IA archaeological sites studied. The pie chart shows the distribution of mitochondrial haplogroups of Uz_IA individuals. (B) PCA of newly sequenced Uz_IA individuals along with published ancient populations from West and East Eurasia projected onto present-day populations. (C) ADMIXTURE ancestry components ($K = 8$) of Uz_IA and published populations from West and East Eurasia. The three main ancestry components are shown in green, red and violet representing ancestry maximized in Anatolian farmers, Iranian farmers, and Hunter Gatherers from West Siberia, respectively. Only the unrelated individuals having $>20k$ SNPs are used for the analyses. The Dehkan site with only one individual ($<20k$ SNPs) has been discarded from any further analyses. Only a subset of ancient samples, which are spatially and temporally important, is shown and supplementary figures of PCA and ADMIXTURE show all present-day and ancient populations.

Material online), with three major sources of ancestries maximized in Iranian farmers, Anatolian farmers and WSHG (Steppe-related), as well as small amounts of West European Hunter Gatherer and East Asian-related ancestries (fig 1C and supplementary fig. S5, Supplementary Material online). The single individual from the Serkharakat site also shows high genetic similarity to Rabat individuals in PCA (fig. 1B and supplementary figs. S1–S3, Supplementary Material online), clustering with other Rabat individuals, and having similar admixture components in ADMIXTURE analysis (fig. 1C).

We additionally investigated the genetic affinity between Rabat and Serkharakat using the outgroup- f_3 test, that is, $f_3(\text{Rabat, Serkharakat/Ancient Population; Mbuti})$ and f_4 -statistics test of the form $f_4(\text{Rabat, Ancient Population; Serkharakat, Mbuti})$ (fig. 2B and supplementary fig. S6, Supplementary Material online). We noticed that Rabat and Serkharakat share strong genetic drift with each other, but both share greater genetic drift with Steppe MLBA and Afanasievo populations, indicating that the Serkharakat individual is not descended from Rabat individuals. In most cases, Uz_IA shows higher genetic affinity to the Late Bronze Age (LBA) populations with Steppe-related ancestry from Turan, Altai, and eastern European regions than to the other BA populations from Turan and Central Asia (fig. 2A and supplementary figs. S6, S8, and S9 and table S3, Supplementary Material online). This trend of higher Steppe-related ancestry in the Uz_IA samples is also supported by their closer proximity to present-day Europeans in the PCA, as well as higher values for $f_3(\text{Uz_IA, Present-day Population; Mbuti}) > 0$, relative to Central Asian (including Turan) and Caucasus populations (supplementary figs. S3 and S7, Supplementary Material online).

The outgroup- f_3 analysis shows the highest genetic affinity between Uz_IA and the Steppe_MLBA population (fig. 2A and supplementary fig. S8, Supplementary Material online). Among the Early-to-Middle Bronze Age Steppe (Steppe_EMBA) populations, Uz_IA shows the highest affinity with the Afanasievo population in outgroup- f_3 statistics and an overall greater affinity to Western_Steppe_EMBA populations than to Central_Steppe_EMBA populations (supplementary fig. S8, Supplementary Material online). Among the Steppe_MLBA populations, Uz_IA shows higher genetic affinity to Central_Steppe_MLBA than to Western_Steppe_MLBA populations (supplementary fig. S8, Supplementary Material online). The f_4 -statistics of the form $f_4(\text{Ancient Population, Ancient Population; Uz_IA, Mbuti})$ also suggest a greater affinity between Steppe MLBA populations and Uz_IA compared with the BMAC populations and other ancient BA populations from the Steppe region except for the Afanasievo population (supplementary fig. S9, Supplementary Material online). We conclude that Uz_IA populations were thus derived from an LBA/IA Steppe pastoralist group that admixed with later IA populations in eastern Asia and Turan, principally derived from populations related to the Central_Steppe_MLBA. We observe a greater genetic affinity of Uz_IA to present-day Europeans than to the present-day Uzbekistan populations (supplementary fig.

S7, Supplementary Material online). This higher genetic affinity for European populations is due to the similar components of Anatolian farmer and Steppe-related ancestries observed both in Uz_IA and European present-day populations. Lower genetic affinity for the present-day Uzbekistan populations indicates substantial demographic changes through several admixture events over the past ~2,000 years whereby present-day Uzbekistan populations now show additional ancestries derived from East Asian and Siberian populations (Irwin et al. 2010; Yunusbayev et al. 2015). Considering the greater inflow of Steppe ancestry during the IA into this region, the present-day populations reveal an interruption of genetic continuity since the Iron Age in southern Uzbekistan.

We then modeled the potential ancestry components using qpAdm to further investigate the admixture amount and ancestral sources in the Uz_IA individuals (Patterson et al. 2012). This admixture modeling was done in two stages: the first stage involved distal modeling, which included ancient pre-Copper Age populations as sources of admixture, and the second stage used proximal modeling where we focused on BA and IA populations as sources. The distal models included up to five sources, and we identified three major ancestries related to Iranian farmers (~31–39%), Anatolian farmers (~30–34%), and Steppe-related ancestry (WSHG) (~15–17%); and three minor ancestries of West European Hunter Gatherer (~7–12%), East Asian (~5–7%), and South Asian Hunter Gatherer represented by Onge population (~8%) (table 1). Both Rabat and Serkharakat show similar models except that the Serkharakat individual also shows affinity for South Asian Hunter Gatherers represented by Onge population.

For the proximal model, we first performed admixture- f_3 tests, a three-population test whereby a significantly negative value indicates the target is a mixture of ancestry related to the other two populations. We found that admixture- $f_3(\text{Population1, Population2; Uz_IA}) < 0$ ($z < -3.0$), where the test population includes temporally and spatially preselected ancient populations used for proximal qpAdm modeling. The top 25 negative admixture- f_3 tests predominantly show a pattern where the Uz_IA population appears to be a mixture of ancestry present in BMAC and MLBA populations (supplementary fig. S14, Supplementary Material online). In proximal source modeling, Uz_IA can be modeled as single source with Ksirov_H_Kushan, a population present within a similar geographical region and time period, indicating a similar ancestry profile as that found in Uz_IA individuals (supplementary table S4, Supplementary Material online). A previous genomic study found that the Kushan individuals (30–380 CE) from South Tajikistan possessed a mixed ancestry of LBA Steppe herders (~43.5%), BMAC populations (~16.8%), and Anatolian farmers (~39.7%) (Narasimhan et al. 2019). Some archaeological texts and records from the Ksirov site in Tajikistan have been regarded as a link between the Kushan and the Yuezhi people (Liu 2001; Narasimhan et al. 2019). Our Uz_IA individuals also show similarities with the Ksirov site in Tajikistan and present a similar ancestral profile having an East Asian-related

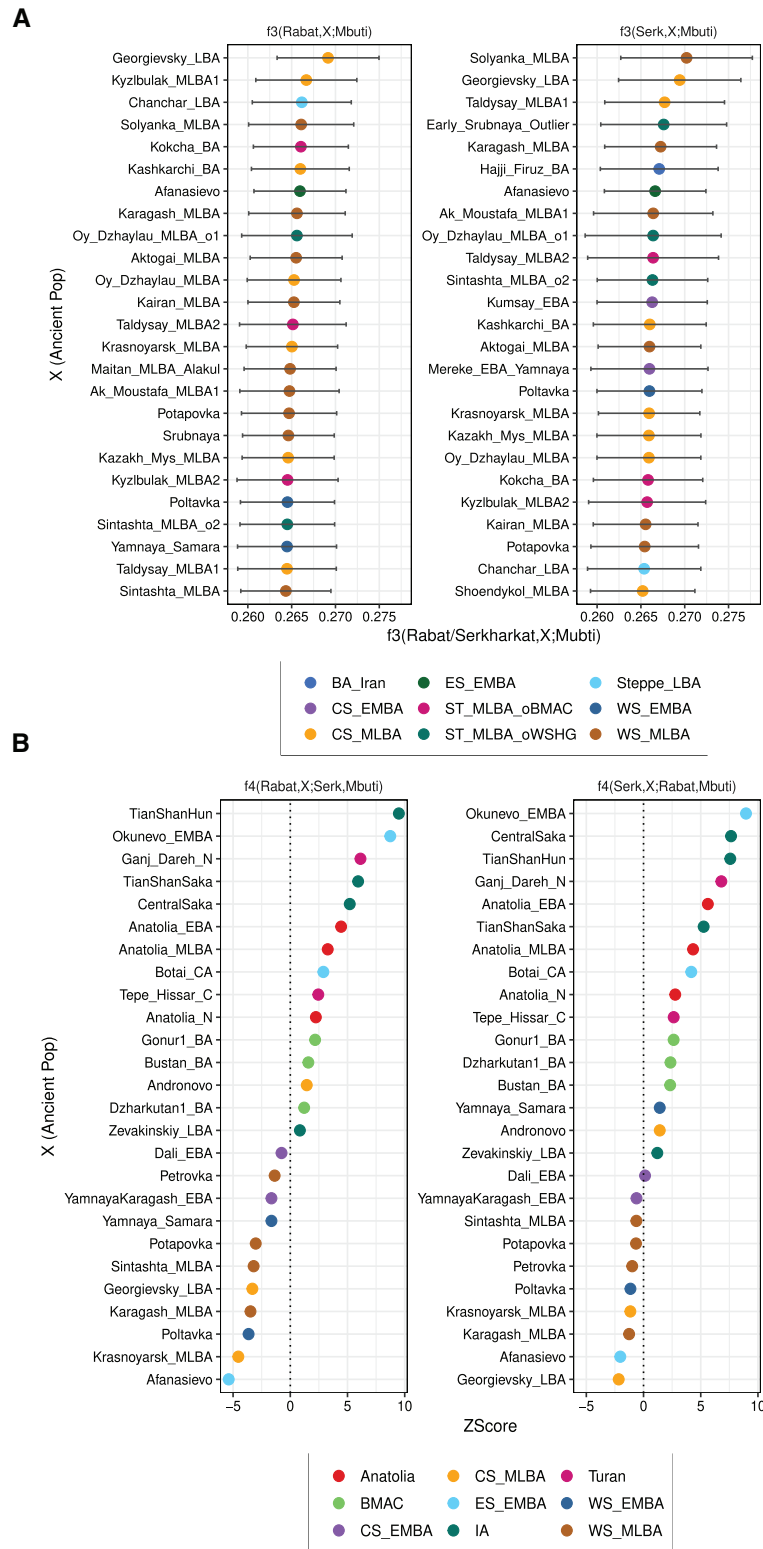


FIG. 2. Outgroup f_3 - and f_4 -statistics of Uz_IA populations. (A) Outgroup- f_3 test ($f_3(Uz_IA, X; Mbuti) > 0$) showing the 25 populations with the highest affinity (most positive) to Uz_IA samples. Most of these populations are from the Central Steppe, and among them MLBA populations from Kazakhstan show the highest affinity. Horizontal bars represent two standard errors. (B) Results of f_4 -statistics of the form $f_4(Rabat/Serkharakat, X; Rabat/Serkharakat, Mbuti)$. The plot shows the significant Z-scores of the f_4 -statistics test results. The negative values show a greater affinity of both Rabat and Serkharakat to the ancient populations compared with each other. The ancient populations are subgrouped as BMAC, Central Steppe EMBA (CS_EMBA), Central Steppe MLBA (CS_MLBA), Eastern Steppe EMBA (ES_EMBA), Steppe LBA (Steppe_LBA), Western Steppe EMBA (WS_EMBA), Western Steppe MLBA (WS_MLBA), Steppe MLBA with additional affinity to BMAC (ST_MLBA_oBMAC), and Steppe MLBA with additional affinity to West Siberian Hunter Gatherer (ST_MLBA_oWSHG). The individuals are grouped similar to Narasimhan et al. (2019).

Table 1. Feasible Five-Way qpAdm Distal Models for Rabat and Serkharakat as Target Population.

Uz_IA	Tail	Source1	Prop1	Source2	Prop2	Source3	Prop3	Source4	Prop4	Source5	Prop5
Rabat	0.304	Iron_Gates_HG	0.073	Ganj_Dareh_N	0.39	Anatolia_N	0.323	West_Siberia_N	0.167	Han	0.047
Rabat	0.255	Iron_Gates_HG	0.071	Ganj_Dareh_N	0.361	Anatolia_N	0.336	West_Siberia_N	0.162	Shamanka_EN	0.07
Serkharakat	0.135	Karelia	0.252	Ganj_Dareh_N	0.376	Anatolia_N	0.32	Shamanka_EN	0.052		
Serkharakat	0.119	Iron_Gates_HG	0.121	Ganj_Dareh_N	0.306	Anatolia_N	0.331	West_Siberia_N	0.157	Onge	0.084
Serkharakat	0.330	Iron_Gates_HG	0.118	Ganj_Dareh_N	0.376	Anatolia_N	0.301	West_Siberia_N	0.151	Shamanka_EN	0.054

NOTE.—The distal models were run for one to five sources, feasible models worked only with four and five sources with a cut-off of $P > 0.05$ and $p_{\text{nest}} < 0.05$ where P value (Tail > 0.5) suggests the admixture from n -source model where “ n ” is the number of source populations and P -nest (< 0.5) indicates the higher ranking n -source model is significantly better than the $n-1$ -source model. In this table, Prop is the admixture proportion and Tail is the P value.

component probably originating from the ancestors of the LBA Steppe pastoralists. Some archaeologists suggest a wider spread for the Yuezhi people and their origins around the northern region of the Tianshan Mountain Region in present-day Northern Xinjiang (Liu 2001), so it may be possible that the Yuezhi people were also derived from the ancestors of Steppe LBA populations with some East Asian ancestry, but additional sampling is needed within this region to fully explore the Yuezhi-related migration and settlement in Bactrian region. Although Kushan period coins were not found in the archaeological record, making it difficult to determine whether this population was directly related to the Kushan Empire that controlled the region during this time period. In a two-way admixture model, Uz_IA can be modeled as a mixture of a Copper Age population (Seh_Gabbi_Iran_Calc) and a Steppe MLBA-related (Georgievsky2_LBA, Priobrazhenka_LBA, and Zevakinskiy_MLBA) ancestry as well as a Ksirov_H_Kushan-related ancestry (supplementary table S4, Supplementary Material online). Finally, using the three proximal sources, many working admixture models reflected a general admixture of Steppe MLBA/LBA-, BMAC-, and Iranian_Chalcolithic-related ancestry. Similar admixture models were also attained for the previously published Kushan individuals (Narasimhan et al. 2019).

In Central Asia, Steppe-related ancestry arrived in the BMAC region by ~4100 BP (de Barros Damgaard, Martiniano, et al. 2018; Narasimhan et al. 2019) and an increased Steppe-related ancestry in Uz_IA populations compared with those dating to the BA supports an increase of contact and admixture with groups from the northern and eastern Steppe of Central Asia. This increase in Steppe-related ancestry in Uz_IA is also observed in other nomadic groups of the Central Steppe such as the Scythians and Sakas and even in the Xiongnu, which were widely present in the Eastern Steppe (de Barros Damgaard, Marchi, et al. 2018). Presence of BMAC-related ancestries further indicate an interaction of Saka- and BMAC-related people which suggests that this mixed ancestry is not only present in Uzbekistan but also in large parts of the Central Steppe region (de Barros Damgaard, Marchi, et al. 2018; Narasimhan et al. 2019). Thus, Uz_IA populations in South Uzbekistan are derived from an LBA Steppe pastoralist population that admixed with a local IA population after the BMAC period. This ancestry was also quite prominent in the late IA Steppe region indicating a high degree of interaction and movement of people in this region.

Genetic Continuity of Bronze Age Ancestry in Iron Age Uzbekistan with Low East and South Asian Connection

Compared with the Bronze Age, Iron Age Central Asia shows increased migrations and cultural exchanges supplemented by technological advances (de Barros Damgaard, Marchi, et al. 2018; de Barros Damgaard, Martiniano, et al. 2018; Narasimhan et al. 2019). Despite major migrations of Steppe-related people into this region, two major components of BA populations, Anatolian and Iranian farmer-related ancestries, still form the majority of Uz_IA ancestry (~64–70%) in ADMIXTURE and qpAdm analyses (fig. 1C and table 1). The two MLBA populations from Uzbekistan (Kashkarchi_BA, Kokcha_BA) and other Steppe_MLBA populations share a closer relationship with Uz_IA than with the BMAC and Indus periphery, Gonur2_BA populations, as shown with f_4 -statistics analysis (supplementary table S3, Supplementary Material online). These MLBA populations have significantly higher Steppe-related ancestry, suggesting a greater interaction with Steppe groups from the Central Steppe region of Kazakhstan in the MLBA period. Thus, a probable genetic connection from the north of Uzbekistan (e.g., Kazakhstan) is evident, bringing Steppe-related ancestry down into this region. Overall, similar core genetic profiles within Uzbekistan from the BA to the IA suggest continuation of BA ancestry into the IA with no mass replacement, although with an increased influx of Steppe-related ancestry, likely from the north of Uzbekistan, possibly using the Inner Asian Mountain Corridor route.

The ADMIXTURE and qpAdm analyses of UZ_IA suggest two major ancestral components related to Iranian and Anatolian farmers (fig. 1C and table 1). These two components also show significantly positive f_4 -statistics ($f_4(\text{Rabat/Serkharakat}, X; \text{Anatolia/Iran}_N, \text{Mbuti})$) compared with the BMAC populations (supplementary fig. S10, Supplementary Material online). Among these two ancestries, we observe a relatively higher Anatolian farmer-related ancestry (~30–34%) in Uz_IA individuals than that found in Neolithic and BA populations in Turan (Bustan_EN: ~9% and BMAC: ~26%), and a lower amount of Iranian farmer-related ancestry in Uz_IA (~31–39%) than found in Bustan_EN: ~73–85% and BMAC ~59% populations (Narasimhan et al. 2019). In proximal qpAdm modeling we also observed that the two- and three-source models always required an Anatolian farmer-related ancestry source (e.g., Hajji_Firuz_C: ~53%), suggesting an increase in Anatolian farmer-related ancestry

in Uz_IA compared with the BMAC populations (supplementary table S2, Supplementary Material online). In addition, we find an increase in Iranian farmer-related ancestry compared with the LBA Steppe populations similar to the other Kushan period individuals (Narasimhan et al. 2019), implying later admixtures with the IA populations from Turan and Iran. This region was traversed extensively by people across Inner Asian Mountain Corridor and the diverse ancestries may also be the result of these movements of people across Central Asia (Frachetti 2012).

Some of the Central Asian populations also show higher affinities with East and South Asian populations apart from the Steppe, Anatolian, and Iranian farmers-related populations (Narasimhan et al. 2019). In the f_4 -statistics test, the Uz_IA populations share more alleles with ancient East Asian populations than with BMAC populations (supplementary fig. S11, Supplementary Material online). This pattern is also observed in both ADMIXTURE and qpAdm analyses (fig 1C and table 1), with the contribution of East Asian-related ancestry ~5–7% (table 1). Additionally, the ADMIXTURE (fig. 1B) and qpAdm models both suggest an ancestry component derived from South Asian Hunter Gatherer ancestry (Onge) (table 1 and supplementary table S4, Supplementary Material online). Between Rabat and Serkharakat, the distal qpAdm model suggests more South Asian Hunter Gatherer ancestry in Serkharakat compared with the Rabat individuals (table 1). In comparison with BMAC populations which contained ~2% ancestry being derived from Onge, an $f_4(\text{Rabat/Serk}; \text{Gonur2_BA/Saidu_Sharif_H}, \text{Mbuti})$ statistics test of Uz_IA with other BA populations shows more shared drift with Onge for the Serkharakat compared with the Rabat (supplementary fig. S13, Supplementary Material online). Therefore, we find relatively more East Asian ancestry in Uz_IA compared with the BMAC populations and relatively similar amounts of South Asian HG ancestry. Additional sampling across this region would provide more clarity for South and East Asian genetic affinities.

Conclusions

In this study, by using the genomic data of 27 Southern Uzbekistan individuals including 15 unrelated individuals from the late IA around the Kushan time period, we report an increase in Steppe-related admixture in the IA compared with BA populations in the Bactria region. The source of Steppe-related ancestry was likely identified as LBA Steppe populations, who admixed with local LBA populations associated with the earlier BMAC populations prevalent in this region. However, the migration and admixture with Steppe-related sources did not replace the previously present Iranian and Anatolian farmer-related ancestry; in fact, we find genetic continuation of BA ancestry with the addition of Steppe-related ancestry. The Uz_IA population also shows a minor genetic connection with populations in South and East Asia, suggesting movement of these ancestries into South Uzbekistan. In addition, increased Anatolian farmer-related ancestry compared with the BA points to additional

admixture with local populations having Anatolian farmer related-ancestry after the decline of the BMAC settlements. This mixed-genetic ancestry had a widespread presence as it is also observed in Kushan-related populations and other Central and Eastern Steppe populations. Future genomic studies in Central Asia will provide more insights into the Iron Age genetic and cultural diversity of this region.

Materials and Methods

Ancient DNA Extraction, Sequencing, and Data Processing

We collected 27 bone remains from three sites in Uzbekistan (supplementary text, Supplementary Material online and fig. 1A).

The detailed archaeological information for all the individuals is provided in the supplementary material, Supplementary Material online. All new samples in this study were provided by the coauthors with full permission from the relevant archaeological institutes or universities and approval for genomic study by the review board of the Institute of Vertebrate Paleontology and Paleoanthropology (Review No. 202101240007). DNA was extracted (Dabney et al. 2013) from bone powder from these remains, and double-stranded libraries were prepared according to published protocols in clean rooms dedicated to ancient DNA preparation (Meyer et al. 2012; Gansauge and Meyer 2013). Mitochondrial sequences were captured using human mitochondrial oligonucleotide probes (Fu et al. 2013) and nuclear DNA was captured using the 1.2 million SNP panel (Fu et al. 2015). Enriched mitochondrial DNA libraries were sequenced on an Illumina Miseq with 2×76 base paired-end reads, and the enriched nuclear DNA libraries were sequenced using the Illumina Hiseq4000 generating 2×76 base paired-end reads.

All reads were trimmed of adapter sequences and paired-end reads with a minimum overlap of 11 bp were merged using leeHom (<https://github.com/grenaud/leeHom>, last accessed June 2020) (Renaud et al. 2014). Merged reads with a minimum length of 30 bp were then mapped to the published human mitochondrial sequence (Andrews et al. 1999) and the hg19 human reference genome using BWA (-aln -n 0.01 and -l 16500). Duplicate reads and those with a mapping quality less than 30 were discarded. Haplogroups for mitochondrial genomes were determined using the Haplogrep2 (Weissensteiner et al. 2016). All nuclear data were genotyped by randomly sampling one read covering each position.

Test for Contamination

The modern DNA contamination rate for each individual sample was estimated using ContamMix (Fu et al. 2016) by aligning mtDNA fragments to the consensus mitochondrial genome and 311 present-day world-wide mtDNA sequences. We ignored the first and last five nucleotides from each fragment, and if >3% of mtDNA fragments matched any of the sampled present-day individuals compared with the consensus sequence, that library was treated as contaminated. All

mtDNA from the individuals analysed had a limited amount of contamination ranging from 0% to 2.7%.

Test for Kinship

The program Relationship Estimation from Ancient DNA (READ) is used to detect the Kinship relationships between Uz_IA individuals (Kuhn et al. 2018). To reliably calculate the kinship relationships, we selected individuals having >20k SNPs and merged with the present-day Human populations also used for PCA (supplementary fig. S3, Supplementary Material online) and pruned the data (–maf 0.05 –indep-pairwise 200 25 0.5) in PLINK to use only the SNPs which are independent of each other (Purcell et al. 2007). From each kinship group, we selected the individual with the higher number of SNPs for subsequent genetic analyses (e.g., f_3 , f_4 -statistics and qpAdm). In total, we discarded three individual samples due to kin relationships L8005, L8619, and L8629 (supplementary table S2, Supplementary Material online).

Mitochondrial and Y-Haplogroup Assignment

All the mtDNA haplogroups of Uz_IA were assigned using the Haplogrep 2 (<https://haplogrep.i-med.ac.at>, last accessed June, 2020) (Weissensteiner et al. 2016). For a comparative mitochondrial analysis with ancient and present-day human populations, we used the published data presented in Irwin et al. (2010) and Wang et al. (2021). The mitochondrial genomes were aligned using MAFFT (Katoh et al. 2002) and F_{st} -distance was determined by Arlequin (Excoffier and Lischer 2010). The male individuals were assigned to different Y-chromosomal haplogroups using the phylogenetic tree in the ISOGG database version 10.01 (<http://www.isogg.org/tree>) with the criteria of using the most derived allele upstream and most ancestral allele downstream of the phylogenetic tree. When substitutions suggestive of ancient DNA damage (C→T or G→A) (Briggs and Heyn 2012) occurred upstream of the most derived SNP, at least two derived SNPs were required for Y-chromosome haplogroup assignment.

PCA and ADMIXTURE

The PCA was carried out using the smartpca program of the EIGENSOFT package (Patterson et al. 2006) with default options and the additional options of Isqproject: YES. We used published ancient populations from Eurasia (Olalde et al. 2014; Raghavan et al. 2014; Skoglund et al. 2014; Allentoft et al. 2015; Haak et al. 2015; Jones et al. 2015; Mathieson et al. 2015; Fu et al. 2016; Jeong et al. 2016; Lazaridis et al. 2016; Yang et al. 2017; de Barros Damgaard, Marchi, et al. 2018; de Barros Damgaard, Martiniano, et al. 2018; Mathieson et al. 2018; Narasimhan et al. 2019; Ning et al. 2019) and present-day populations including Human Origins data set for the 1240k panel downloaded from <https://reich.hms.harvard.edu/>, last accessed June 2020 (Patterson et al. 2012; Lazaridis et al. 2014; Raghavan et al. 2014; Lazaridis et al. 2016; Mallick et al. 2016; Zhang et al. 2017; de Barros Damgaard, Martiniano, et al. 2018) and merged our Uz_IA individuals. In total, we used 67 recent populations (supplementary fig. S3, Supplementary Material online) and 90 ancient populations (supplementary fig. S1, Supplementary

Material online) and projected all the ancient individuals onto the PCA capturing the genetic variation of all the present-day individuals to carry out the PCA. The ADMIXTURE analysis was performed after pruning for linkage disequilibrium using PLINK (v1.3.0) with parameters –maf 0.05 and –indep-pairwise 200 25 0.5 (Alexander et al. 2009).

f_3 - and f_4 -Statistics

Outgroup f_3 -statistics were carried out using the ADMIXTOOLS (Patterson et al. 2012) program qp3Pop, of the form $f_3(\text{Mbuti}; X, Y)$, where the outgroup is the central African Mbuti and X and Y include the diverse ancient populations from Eurasia. In addition, we also calculated the admixture f_3 -statistics, $f_3(\text{Population1, Population2}; \text{Uz_IA}) < 0$ with inbreed: YES parameter for admixture scenario in Uz_IA population with the preselected populations used in the qpAdm analyses. We also performed f_4 -statistics of the form $f_4(X, Y; Z, \text{Mbuti})$, with Mbuti as an outgroup. Significantly positive f_4 -statistics suggest that Z shares more alleles with X than Y, whereas significantly negative f_4 -statistics suggest that Z shares more alleles with Y than X.

qpAdm Analyses

The qpAdm software implemented in ADMIXTOOLS (Patterson et al. 2012) was used for admixture modeling in two stages. In the first stage, distal modeling was done where we used pre-Copper age populations as potential sources for the Uz_IA. In the second stage, we used proximal modeling, where we searched for populations that were temporally and geographically located close to the target populations. In distal modeling, we used the following outgroups:

Mbuti, UstIshim, Kostenki14, Karelia, Malta1, GoyetQ116-1_N, Vestonice16, Villabruna, ElMiron, Natufian_HG, Tianyuan, Ganj_Dareh_N, Anatolia N, West Siberia N, Onge, Han, Shamanka_EN, Iron_Gates_HG.

The populations shown in the italics were used as a potential sources for the Uz_IA populations using a similar strategy to that used in Narasimhan et al. (2019). In proximal modeling, we searched for the more recent populations (after the BA) by fixing the Right populations. The outgroups used include:

Mbuti, UstIshim, Kostenki14, Karelia, Malta1, GoyetQ116-1_N, Vestonice16, Villabruna, ElMiron, Natufian_HG, Ganj_Dareh_N, Anatolia_N, West_Siberia_N, Onge, Tianyuan, Shamanka_EN, Iron_Gates_HG, Han. The list of source populations used for proximal modeling is given in supplementary table S5, Supplementary Material online. In both distal and proximal modeling, we selected models with $P > 0.05$ and $p_{\text{nest}} < 0.05$ for the tested sources.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank archaeological teams from Northwest University, Xi'an. This work was supported by National Key R&D Program of China (2016YFE0203700), the Chinese

Academy of Sciences (CAS, XDB26000000). National Natural Science Foundation of China (91731303, 41925009, 41630102), Chinese Academy of Sciences (XDA1905010, QYZDB-SSW-DQC003), Tencent Foundation through the XPLOER PRIZE, and the Howard Hughes Medical Institute (Grant No. 55008731).

Author Contributions

Q.F. designed the research project. Q.F. and V. K. managed the project. J.W., D.Z., Y.L., Y. T., M.R., B.A., H.M., M.Z., M.D. assembled archaeological materials and dating. P.C., R.Y., F.L., X.F., Q.D., W.P. performed or supervised wet laboratory work. Q.F. and X.F. did the data processing and quality control. V.K. analysed the data and interpreted the results. V.K. and Q.F. wrote the manuscript with revision from E.A.B. and M.A.Y. All authors discussed, critically revised, and approved the final version of the manuscript.

Data Availability

All newly sequenced individuals mapped BAM, genotype calls, and mitochondrial fasta files (Wang et al. 2017) are available and deposited in BIG Data Center (Big Data Center Members 2018) (<https://bigd.big.ac.cn/gsa-human>; accession number: PRJCA004828).

References

- Abazov R. 2008. The Palgrave concise historical atlas of Central Asia. New York: Palgrave Macmillan.
- Abdullaev K. 2007. Nomad migration in Central Asia. *Proc Brit Acad.* 133:73–98.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167–172.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 23(2):147.
- Big Data Center Members. 2018. Database resources of the BIG Data Center in 2018. *Nucleic Acids Res.* 46(D1):D14–D20.
- Briggs AW, Heyn P. 2012. Preparation of next-generation sequencing libraries from damaged DNA. *Methods Mol Biol.* 840:143–154.
- Cunliffe B. 2015. By steppe, desert, and ocean: the Birth of Eurasia. Oxford: Oxford University Press.
- Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga J-L, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A.* 110(39):15758–15763.
- de Barros Damgaard P, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliusson T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E, et al. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557(7705):369–374.
- de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, 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(6396):eaar7711.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10(3):564–567.
- Frachetti MD. 2012. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr Anthropol.* 53(1):2–38.
- Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524(7564):216–219.
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, Pääbo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A.* 110(6):2223–2227.
- 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(7606):200–205.
- Gansauge M-T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protocols.* 8:737–748.
- 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(7555):207–211.
- Irwin JA, Ikramov A, Saunier J, Bodner M, Amory S, Röck A, O’Callaghan J, Nuritdinov A, Atakhodjaev S, Mukhamedov R, et al. 2010. The mtDNA composition of Uzbekistan: a microcosm of Central Asian patterns. *Int J Legal Med.* 124(3):195–204.
- Jeong C, Ozga AT, Witonsky DB, Malmström H, Edlund H, Hofman CA, Hagan RW, Jakobsson M, Lewis CM, Aldenderfer MS, 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 U S A.* 113(27):7485–7490.
- Jones ER, Gonzalez-Forbes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun.* 6:8912.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–3066.
- Kohl P. 2007. The making of Bronze Age Eurasia. Cambridge: Cambridge University Press.
- Kuhn JMM, Jakobsson M, Günther T. 2018. Estimating genetic kin relationships in prehistoric populations. *PLoS One.* 13(4):e0195491.
- Kuhr A. 2001. The Achaemenid Persian Empire (c.550–c.330 BCE): continuities, adaptations, transformations. In: Alcock S, D’Alroy T, Morrison KD, Sinopoli CM, editors. Empires: perspectives from archaeology and history. Cambridge: Cambridge University Press. p. 92–124.
- Lazaridis I, Nadel D, Rollefson G, Merrett DC, 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(7617):419–424.
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.
- Lhuillier J, Mashkour M. 2017. Animal exploitation in the oases: an archaeological review of Iron Age sites in southern Central Asia. *Antiquity* 91(357):655–673.
- Liu X. 2001. Migration and settlement of the Yuezhi-Kushan: interaction and interdependence of nomadic and sedentary societies. *J World Hist.* 12(2):261–292.
- 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(7624):201–206.
- Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkoshbacht N, Candilio F, Cheronet O, et al. 2018. The genomic history of southeastern Europe. *Nature* 555(7695):197–203.

- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503.
- Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222–226.
- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, et al. 2019. The formation of human populations in South and Central Asia. *Science* 365(6457):eaat7487.
- Ning C, Wang C-C, Gao S, Yang Y, Zhang X, Wu X, Zhang F, Nie Z, Tang Y, Robbeets M, et al. 2019. Ancient genomes reveal Yamnaya-related ancestry and a potential source of Indo-European speakers in Iron Age Tianshan. *Curr Biol*. 29(15):2526–2532.e4.
- Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M, Prado-Martínez J, Rodríguez JA, Rasmussen S, Quilez J, et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old mesolithic European. *Nature* 507(7491):225–228.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet*. 2(12):e190.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81(3):559–575.
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Orlando L, Metspalu E, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.
- Renaud G, Stenzel U, Kelso J. 2014. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res*. 42(18):e141.
- Sarmiento J, Lhuillier J. 2011. Iron Age in Turkmenistan: Ulug depe in the Kopetdagh piedmont. M. Mamedow. Historical and cultural sites of Turkmenistan. Discoveries, researches and restoration for 20 years of independence. Turkmen State Publishing Service. Halshs-01111871. Available from: <https://halshs.archives-ouvertes.fr/halshs-01111871>.
- Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, Hall P, Tambets K, Parik J, Sjögren K-G, et al. 2014. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science*. 344(6185):747–750.
- Spengler R. 2015. Agriculture in the Central Asian Bronze Age. *J World Prehist*. 28(3):215–253.
- Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, et al. 2015. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet*. 23(1):124–131.
- Wang W, Ding M, Gardner JD, Wang Y, Miao B, Guo W, Wu X, Ruan Q, Yu J, Hu X, et al. 2021. Ancient Xinjiang mitogenomes reveal intense admixture with high genetic diversity. *Biochem Pharmacol*. 24(17):eabd6690–1641.
- Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, Tang B, Dong L, Ding N, Zhang Q, et al. 2017. GSA: genome sequence archive. *Genomics Proteomics Bioinformatics*. 15(1):14–18.
- Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, Schönherr S. 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. 44(W1):W58–W63.
- Wu X, Miller FN, Crabtree P. 2015. Agro-pastoral strategies and food production on the achaemenid frontier in Central Asia: a case study of Kyzyltepa in Southern Uzbekistan. *J Brit Inst Persian Stud*. 53:93–117.
- Yang MA, Gao X, Theunert C, Tong H, Aximu-Petri A, Nickel B, Slatkin M, Meyer M, Pääbo S, Kelso J, et al. 2017. 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr Biol*. 27(20):3202–3208.e9.
- Yunusbayev B, Metspalu M, Metspalu E, Valeev A, Litvinov S, Valiev R, Akhmetova V, Balanovska E, Balanovsky O, Turdikulova S, et al. 2015. The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet*. 11(4):e1005068.
- Zhang C, Lu Y, Feng Q, Wang X, Lou H, Liu J, Ning Z, Yuan K, Wang Y, Zhou Y, et al. 2017. Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biol*. 18(1):115.