

EVOLUTIONARY BIOLOGY

Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians

Guido Alberto Gnecci-Ruscione^{1,2}, Elmira Khussainova³, Nurzhibek Kahbatkyzy^{3,4}, Lyazzat Musralina^{1,3,4}, Maria A. Spyrou¹, Raffaella A. Bianco¹, Rita Radzeviciute¹, Nuno Filipe Gomes Martins¹, Caecilia Freund¹, Olzhas Iksan^{3,4}, Alexander Garshin³, Zhassulan Zhaniyazov³, Bakhytzhon Bekmanov^{3,4}, Egor Kitov^{4,5,6}, Zainolla Samashev^{7,8}, Arman Beisenov⁶, Natalia Berezina⁹, Yakov Berezin⁹, András Zsolt Bíró¹⁰, Sándor Évinger¹⁰, Arman Bissembaev^{8,11}, Gaziz Akhatov⁶, Aslan Mamedov¹², Akhan Onggaruly¹³, Dmitriy Voyakin^{6,14}, Aidos Chotbayev⁶, Yeldos Kariyev¹⁵, Alexandra Buzhilova⁹, Leyla Djansugurova^{3*}, Choongwon Jeong^{1,16*}, Johannes Krause^{1,2*}

The Scythians were a multitude of horse-warrior nomad cultures dwelling in the Eurasian steppe during the first millennium BCE. Because of the lack of first-hand written records, little is known about the origins and relations among the different cultures. To address these questions, we produced genome-wide data for 111 ancient individuals retrieved from 39 archaeological sites from the first millennia BCE and CE across the Central Asian Steppe. We uncovered major admixture events in the Late Bronze Age forming the genetic substratum for two main Iron Age gene-pools emerging around the Altai and the Urals respectively. Their demise was mirrored by new genetic turnovers, linked to the spread of the eastern nomad empires in the first centuries CE. Compared to the high genetic heterogeneity of the past, the homogenization of the present-day Kazakhs gene pool is notable, likely a result of 400 years of strict exogamous social rules.

INTRODUCTION

The transition to the Iron Age (IA) marks one of the most important events in the history of Eurasia. At the turn of the first millennium BCE, changes in the archeological record attest to the rise of several nomad cultures across the steppe, from the Altai to the western fringe of the Pontic-Caspian region (1). These cultures are often collectively referred to as Scythians based on the common features found in their mortuary contexts (2). Compared to the preceding Bronze Age (BA) populations, the Scythians went through a transition from a sedentary to a nomadic cattle-breeding lifestyle, showed

an increase in warfare and advancements in military technologies (e.g., new types of iron weapons and horseback riding techniques, such as introducing the use of a saddle), and the establishment of hierarchical elite-based societies (3).

Previous genomic studies have detected large-scale genetic turnovers (and therefore substantial human migrations) in the BA steppe, which eventually resulted in the formation of a homogeneous and widespread Middle and Late BA (LBA) gene pool that characterized the sedentary herders of the western and central steppe (“steppe_MLBA”) (4–7). The reasons that prompted the rapid decline of these MLBA cultures and the rise of the Scythians are still poorly understood. Scholars have pointed out, among the most relevant factors, the climatic humidification (8) and socioeconomic pressures from the neighboring farming civilizations, i.e., the ones linked to the Bactria Margiana Archaeological Complex (“BMAC”) (3). Three competing hypotheses have been debated regarding the origins of the Scythians: a Pontic-Caspian origin, supported by their assumed Iranian languages, a Kazakh Steppe origin supported by the archaeological findings, and a multiple independent origin from genetically distinct groups that adopted common cultural traits (2). The limited number of genomes so far retrieved from the IA steppe nomads provides a glimpse of their genetic diversity but is far from being sufficient to characterize complex patterns of admixture between various eastern and western Eurasian gene pools (9–12).

From an archaeological perspective, the earliest IA burials associated with nomad-warrior cultures were identified in the eastern fringes of the Kazakh Steppe, in Tuva and the Altai region (ninth century BCE) (13). Following this early evidence, the Tasmola culture in central and north Kazakhstan is among the earliest major IA nomad warrior cultures emerging (eighth to sixth century BCE) (13). These earlier groups were followed by the iconic Saka cultures located in southeastern Kazakhstan and the Tian Shan mountains

¹Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. ²Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. ³Laboratory of Population Genetics, Institute of General Genetics and Cytology, 050060, al-Farabi Ave., 93 Almaty, Kazakhstan. ⁴Kazakh National University by al-Farabi, 050040, al-Farabi Ave., 71 Almaty, Kazakhstan. ⁵Center of Human Ecology, Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow 119991, Russian Federation. ⁶Institute of Archaeology by A.Kh. Margulan, 050010, Dostyk Ave., 44 Almaty, Kazakhstan. ⁷State Historical and Cultural Museum-Reserve “Berel,” village Zhambyl, Katon-Karagay district, East Kazakhstan region, 070906, Kazakhstan. ⁸Branch of Institute of Archaeology by A.Kh. Margulan, 010011, Republic Ave., 24, of. 511, Nur-Sultan, Kazakhstan. ⁹Research Institute and Museum of Anthropology, Lomonosov Moscow State University, Mokhovaya Str. 11, Moscow 125009, Russian Federation. ¹⁰Department of Anthropology, Hungarian Natural History Museum, Ludovika tér 2-6, Budapest H-1083, Hungary. ¹¹Regional Center of History, Ethnography and Archeology of Aktobe region, Turgenev Str. 86, Aktobe 030020, Kazakhstan. ¹²Aktobe Regional Historical Museum, Orazta Tateuly Ave. 3, Batsys-2 microdistrict, Astana district, Aktobe 030000, Kazakhstan. ¹³Scientific-Research Institute “Halyq qazynasy,” National Museum of the Republic of Kazakhstan, Tauelsizdik Ave. 54, Nur-Sultan 010000, Kazakhstan. ¹⁴International Institute for Central Asian Studies, University Boulevard Street 9, Samarkand 140129, Uzbekistan. ¹⁵Scientific Center “Altaytanu” of East-Kazakhstan State University by S.A. Amanzholov, Kazakhstan Str. 55, Ust-Kamenogorsk 070004, Kazakhstan. ¹⁶School of Biological Sciences, Seoul National University, 08826 Seoul, Republic of Korea.

*Corresponding author. Email: leylad@mail.ru (L.D.); cwjeong@snu.ac.kr (C.J.); krause@shh.mpg.de (J.K.)

(sixth to second century BCE), the Pazyryk culture centered in the Altai mountains (fifth to first century BCE-CE), and the Sarmatians that first appeared in the southern Ural region (sixth to second century BCE) and later are found westward as far as the northern Caucasus and eastern Europe (fourth BCE to fourth CE) (1, 14–17). The nomad groups also influenced their sedentary neighbors, such as the ones associated with the Sargat cultural horizon (fifth to first century BCE) located in the northern forest-steppe zone between the Tobol and Irtysh rivers (3, 18).

After the IA, the Kazakh Steppe served as a center for the expansion of multiple empires, such as the Xiongnu and Xianbei chiefdoms from the east (19) and the Persian-related kingdoms from the south (e.g., Kangju) (20). These events brought the demise of the eastern Scythian cultures, but the demographic turnovers associated with this cultural transition remain poorly understood (20). Furthermore, forms of nomadic lifestyle persisted in the Kazakh Steppe throughout the centuries. A key event in the recent history of the nomad populations happened in the 15th to 16th century CE when all the tribes living in the territory of present-day Kazakhstan were organized and grouped into three main hordes (Zhuzs): Elder Zhuz, Middle Zhuz, and Junior Zhuz located in southeast, central/northeast, and west Kazakhstan, respectively (21, 22). This division was a political and religious compromise between different nomadic tribes, which were spread across Central Asia and had to protect themselves from external threats after the collapse of the Golden Horde. This set the basis for the foundation of the Kazakh Khanate (1465 to 1847 CE). Today, Kazakh groups in Kazakhstan still maintain their tribal affiliations and revere their nomadic history preserving some aspects of its culture (21). One of these traditions is the “Zheti-ata,” which consists of keeping track of the family tree up to seven generations by paternal line to avoid marriage between kins (23).

To understand the genetic structure of the different IA nomadic cultures as well as the demographic events associated with their origins and decline, we successfully generated genome-wide data from 111 ancient human individuals retrieved from 39 different archaeological sites across the Kazakh Steppe (Kazakhstan, Kyrgyzstan, and Russia) and one individual retrieved from a Hun elite burial located in present-day Hungary (text S1). Our dataset densely covers a time span ranging from the eighth century BCE to the fourth century CE and also includes three individuals from the medieval period (Fig. 1 and text S1). We also produced new genome-wide data from 96 modern-day Kazakh individuals belonging to several tribes affiliated to the three major Kazakh hordes (Zhuzh) covering the entire territory of present-day Kazakhstan to better understand how recent historical events have shaped the genetic structure of present-day nomads.

RESULTS

Genome-wide data for 117 ancient individuals were obtained using an in-solution DNA capture technique designed to enrich for 1,233,013 single-nucleotide polymorphisms (SNPs) commonly referred as 1240K capture (Materials and Methods). Genome-wide data for 96 present-day Kazakh individuals were generated with the Affymetrix Axiom Genome-wide HumanOrigins SNP-chip (“HO”) (Materials and Methods). After performing quality controls, we retained all the 96 modern Kazakh individuals and 111 ancient individuals with at least >20,000 SNPs covered, obtaining a median of 793,636 successfully recovered SNPs and 1.5× autosomal coverage

on the 1240K panel across all individuals (Materials and Methods and data file S1).

We then merged the new data with a reference dataset of previously published modern and ancient individuals compiling a “1240KHO” dataset consisting of 586,594 SNPs overlapping with the modern genotype data that we used for performing global population structure analyses [i.e., PCA (principal components analysis) and ADMIXTURE]. We also produced a “1240K”-only dataset consisting of 1240K capture or whole-genome shotgun data pooled down to include 1240K sites only that we used for the rest of the analyses (Materials and Methods and tables S2 and S3). For population-based analyses, we grouped individuals according to their archaeological culture affiliation, spanning a defined time range after excluding genetic outliers shifted more than ± 2 SD from the median PCs of their respective group (Materials and Methods and table S1).

The IA transition in the Kazakh Steppe

Overall, PCA and ADMIXTURE suggest that a substantial demographic shift occurred during the transition from the BA to the IA in the Kazakh Steppe (Fig. 2 and figs. S1 and S2). In contrast to the highly homogeneous steppe_MLBA cluster found across the Kazakh Steppe until the end of the second millennium BCE, the IA individuals are scattered across the PC space, most notably along PC1 and PC3. Their spread along these PCs suggests a varying degree of extra eastern Eurasian affinity compared to the MLBA population and extra affinity to southern populations ultimately related to the Neolithic Iranians and the Mesolithic Caucasus hunter-gatherers (from here on referred to as Iranian-related ancestry), respectively. Despite the high genetic variability, it is possible to appreciate homogeneous clusters of ancient individuals belonging to the same archaeological culture and/or geographic area (Fig. 2 and fig. S1). Following a chronological order, most of the individuals from the sites associated with the Early IA Tasmola culture (“Tasmola_650BCE”) and the published “Saka_Kazakhstan_600BCE” of central-north Kazakhstan cluster together in the middle of the PCA plot and show a uniform pattern of genetic components in ADMIXTURE analyses (Fig. 2, A and D, and figs. S1 and S2). The two previously published individuals from the Aldy Bel site in Tuva (Aldy_Bel_700BCE) also fall within this genetic cloud (Fig. 2A). This genetic profile persists in the later Middle and Late IA, shown by most individuals from the Pazyryk site of Berel (“Pazyryk_Berel_50BCE”) (Fig. 2B). This IA cluster is distinct from the previous steppe_MLBA groups inhabiting the same regions, most notably because of its substantial shift toward eastern Eurasians along PC1. In addition, we find outliers showing an even stronger shift to eastern Eurasians than the main cluster: two outliers from Pazyryk Berel time (“Pazyryk_Berel_50BCE_o”), three outliers from the Tasmola site of Birlik (“Tasmola_Birlik_640BCE”), and three of four individuals from the Korgantas phase of central-north Kazakhstan (24) (Fig. 2B and table S2). One female individual from Birlik (BIR013.A0101) with an eastern Eurasian genetic profile was unearthed with grave goods (a bronze mirror) that presented typical Eastern Steppe features (text S1).

The classical IA Sakas from the Tian Shan region to the south (“Saka_TianShan_600BCE,” “Saka_TianShan_400BCE,” and previously published “Pub_Saka_TianShan_200BCE”) are distributed along a cline between the Tasmola/Pazyryk cluster and the Iranian-related gene pool, along PC3 (Fig. 2, A and B). A stronger affinity to the

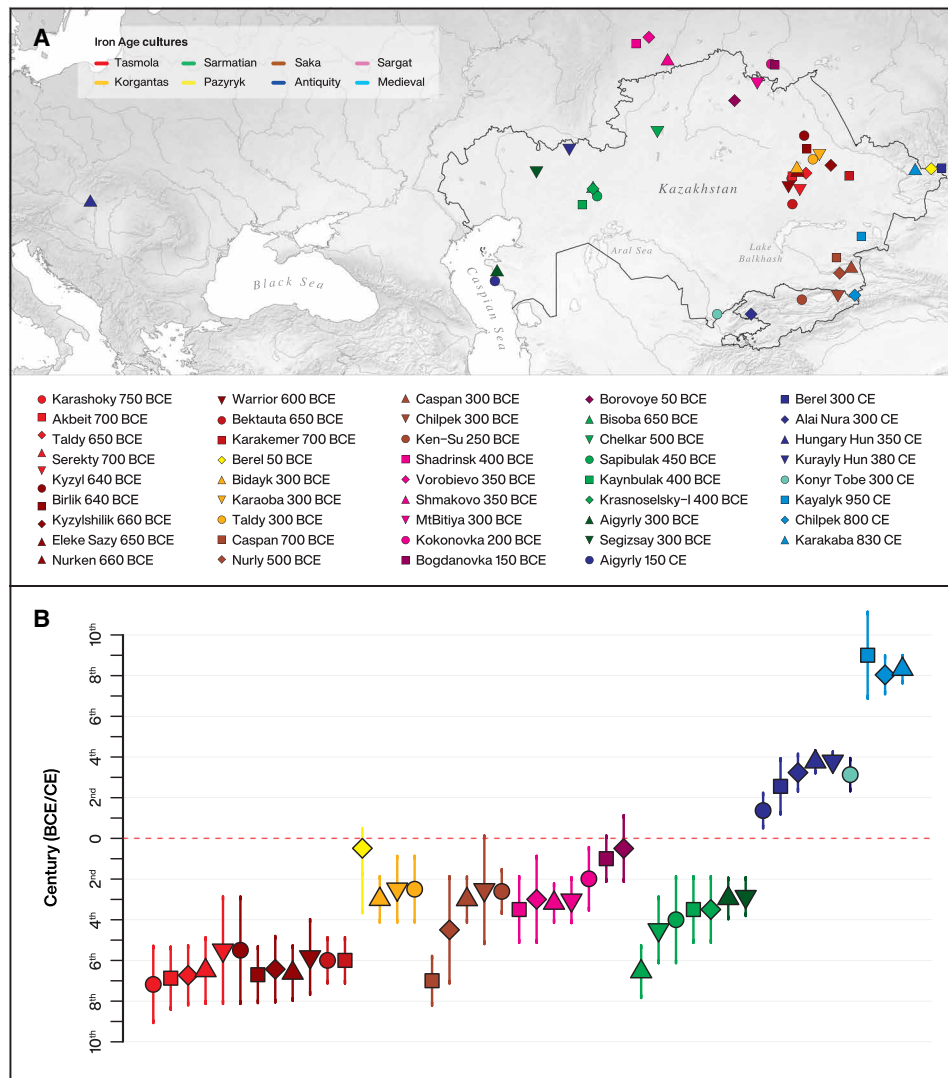


Fig. 1. Geographic location and dates of the newly reported ancient genomes. (A) Map showing the locations of the 39 archaeological sites where the 117 individuals were retrieved and (B) their respective dates in years BCE/CE. The dates reported are ^{14}C -calibrated (2-sigma) ranges for the sites comprehending at least one individual directly radiocarbon-dated; if more individuals are dated, we report the lowest and the highest values across all of them. If for a site, no individuals are dated, we report the date ranges based on the archaeological context (data file S1). The sites are colored according to their cultural affiliation. This same culture-based color code (top right) is maintained for all the figures in the main text and the Supplementary Materials.

Neolithic Iranians is also found in ADMIXTURE analyses (Fig. 2D and fig. S2). The shift toward the Iranian-related gene pool is found as early as ~650 BCE in one Eleke_Sazy_650BCE individual (ESZ002) retrieved from an elite Saka burial, while three of four individuals from one of the earliest Tian Shan Saka site of Caspan_700BCE fall within the Tasmola/Pazyryk cloud.

The individuals associated with the sedentary Sargat culture in the forest-steppe zone north of the Kazakh Steppe (“Sargat_300BCE”) partially overlap with the Tasmola/Pazyryk cluster although forming a cloud in PCA that is shifted toward western Eurasians and toward the uppermost cline of northern Inner Eurasians (PC1 and PC2, respectively; Fig. 2B). In line with PCA, Sargat individuals carry a small proportion of a different type of northeast Asian ancestry not detected in the nomad groups further to the south (Fig. 2D).

With the exception of one outlier falling in the Tasmola/Pazyryk cloud, the individuals associated with the Sarmatian culture are

highly homogeneous despite being spread over a wide geographic area and time period (i.e., early “Sarmatians_450BCE,” late “Sarmatians_150BCE,” and western “Sarmatians_CaspianSteppe_350BCE”; Fig. 2, A and B). Our new data from seven early Sarmatian sites in central and western Kazakhstan (Sarmatians_450BCE) document that this gene pool was already widespread in this region during the early phases of the Sarmatian culture. Furthermore, Sarmatians show a sharp discontinuity from the other IA groups by forming a cluster shifted toward west Eurasians (Fig. 2 and table S2).

Admixture modeling of IA steppe populations

Genetic ancestry modeling of the IA groups performed with qpWave and qpAdm confirmed that the steppe_MLBA groups adequately approximate the western Eurasian ancestry source in IA Scythians while the preceding steppe_EBA (e.g., Yamnaya and Afanasievo) do not (data file S4). As an eastern Eurasian proxy, we chose LBA

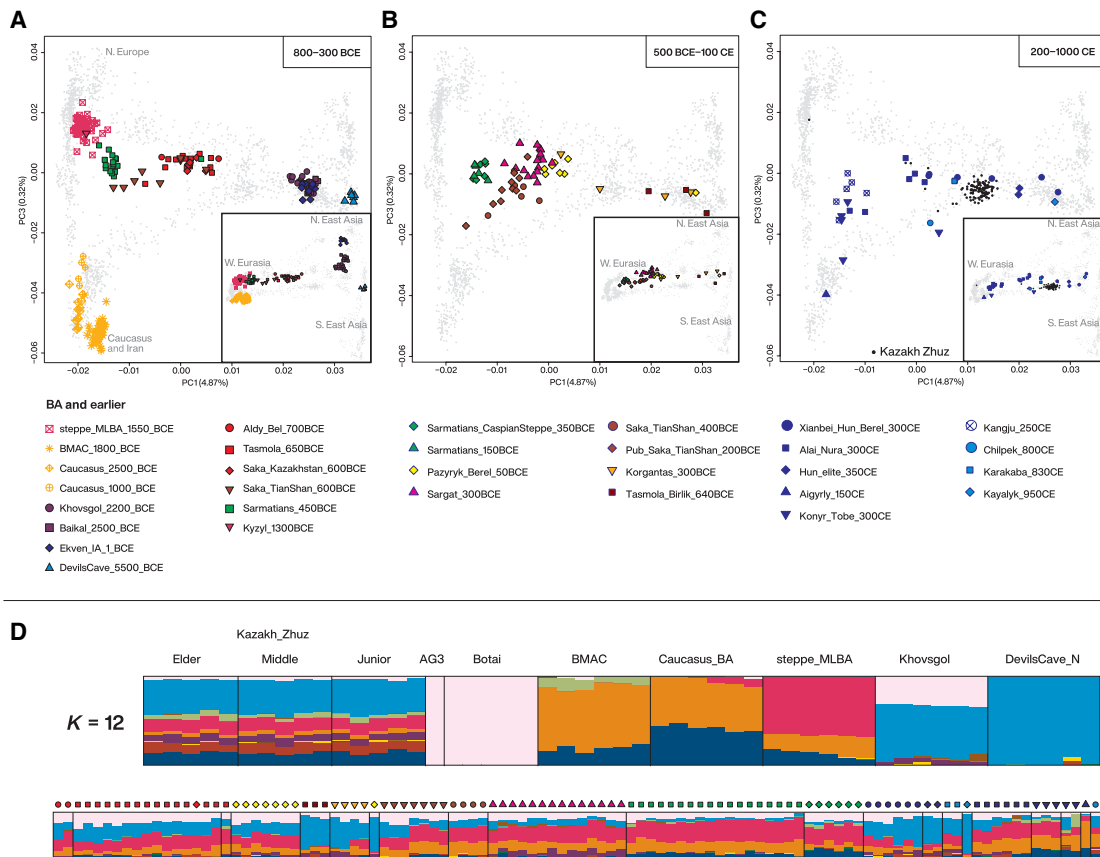


Fig. 2. PCA and ADMIXTURE analyses. (A to C) PC1 versus PC3 (outer plot) and PC1 versus PC2 (inner plot in the bottom right box) including all the IA, new and previously published individuals (filled symbols), relevant published temporally preceding groups (empty symbols), and present-day Kazakh individuals (small black points). The gray labels in this and the following panel indicate broad geographical groupings of the modern individuals used to calculate PCA that in the plots are shown as small gray points. The ancient samples are distributed in (A) to (C) sliced in three different time intervals as reported in the top right corner. (D) Histograms of ADMIXTURE analysis ($K = 12$; fig. S2) for the new IA and post-IA individuals and selected subset of temporally preceding groups maximizing key genetic components and a randomly selected subset of present-day Kazakh from the three main Zhuzs.

herders from Khovsgol in northern Mongolia based on their geographic and temporal proximity. Other eastern proxies fail the model because of a lack or an excess of affinity toward the Ancient North Eurasian (ANE) lineage (25). However, this two-way admixture model of Khovsgol + steppe_MLBA does not fully explain the genetic compositions of the Scythian gene pools (data file S4). We find that the missing piece matches well with a small contribution from a source related to ancient populations living in the southern regions of the Caucasus/Iran or Turan [we use the term “Turan” for consistency with (7), only its geographical meaning, designating the southern part of Central Asia; Fig. 3A]. The proportions of this ancestry increase through time and space: a negligible amount in the most northeastern Aldy_Bel_700BCE group, ~6% in the early Tasmola_650BCE, ~12% in Pazyryk_Berel_50BCE, ~10% in Sargat_300BCE, ~13% in Saka_TianShan_600BCE, and ~20% in Saka_TianShan_400BCE (Fig. 3A), in line with f_4 -statistics (table S2). Sarmatians also require 15 to 20% Iranian ancestry while carrying substantially less Khovsgol and more steppe_MLBA-related ancestry than the eastern Scythian groups.

For Sarmatians and later Tian Shan Sakas, only the groups from Turan (i.e., Turan_ChL, BMAC, and postBMAC) match as sources, while groups from Iran and Caucasus fail; we chose BMAC and

postBMAC as the representative proxies (Fig. 3A and data file S4). The extra eastern Eurasian influx in the outliers (Tasmola_Birlik_640BCE, Korgantas_300BCE, and Pazyryk_Berel_50BCE_o) is not sourced from the same eastern proxies as the previous groups (i.e., Khovsgol); instead, it can only be modeled with an ancient northeast Asian (ANA) lineage, represented by the early Neolithic groups from the Devil’s Gate Cave site in the Russian Far East (DevilsCave_N) (Fig. 3A and data file S4).

Post-IA genetic turnovers in the Kazakh Steppe

We observe an intensification of the new eastern Eurasian influx described above among the individuals from the early 1st millennium CE (“Xianbei_Hun_Berel_300CE”) as well as the later 7th- to 11th-millennium CE individuals (“Karakaba_830CE” and “Kayalyk_950CE”). They are scattered along PC1 from the main IA Tasmola/Pazyryk cluster toward the ANA groups (Fig. 2C). The two individuals associated with Hun elite burials dated from the third century CE, one from the site of Kurayly in the Aktobe region in western Kazakhstan and the other from Budapest, Hungary (“Hun_elite_350CE”), cluster closely together along this cline (Fig. 2C and figs. S1 to S3).

The individuals from the ancient city of Otyrar Oasis in southern Kazakhstan show a quite distinct genetic profile. Three of five

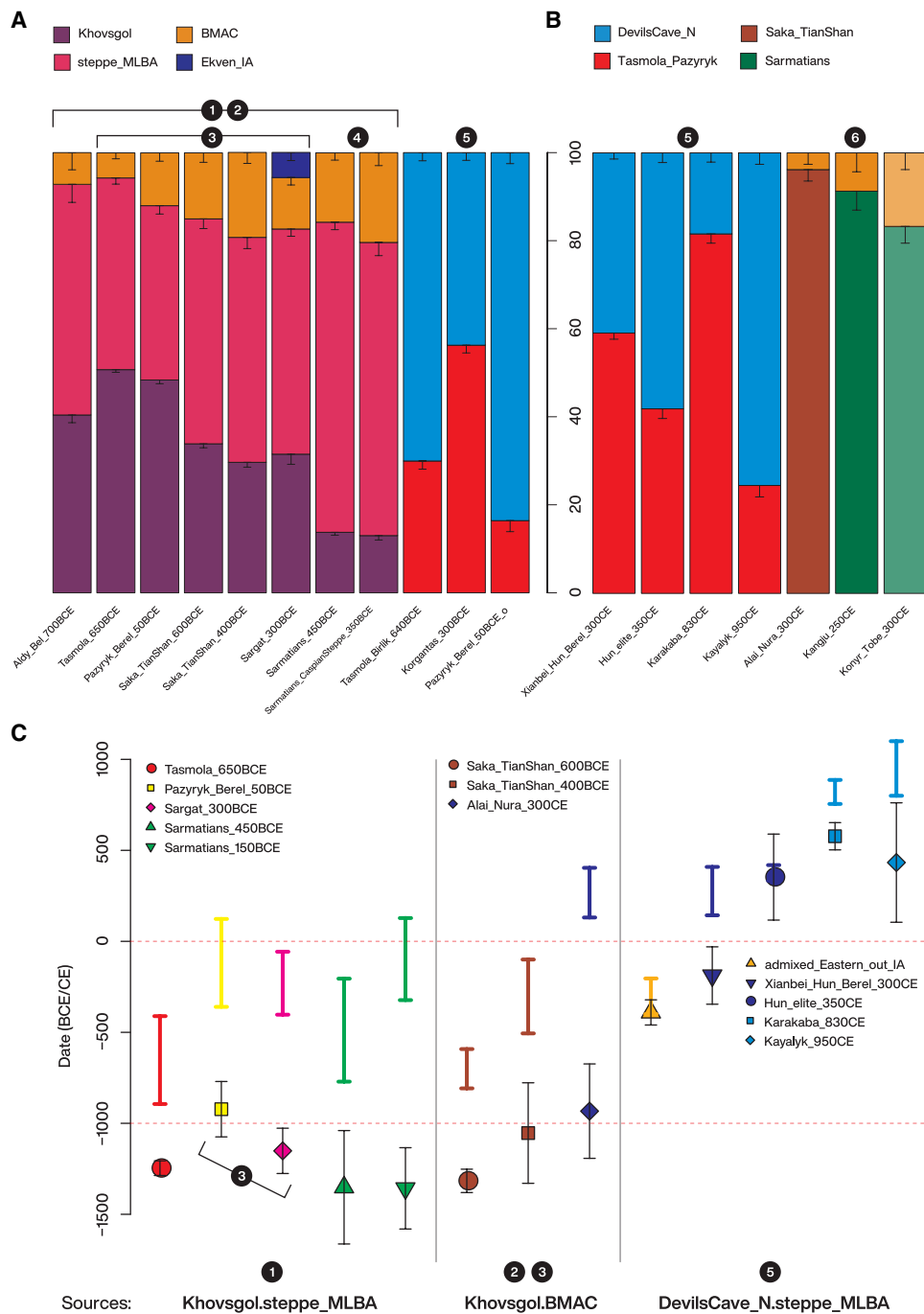


Fig. 3. Bar plots showing the ancestry proportions and SEs obtained from qpWave/qpAdm modelings. (A) Fitting models for the main IA groups using LBA sources, the major genetic shift with the “new” East Asian influx (DevlisCave_N-like) observed in the Middle IA outliers and Korgantas. (B) Fitting models for the post-IA groups using IA groups as sources. A transparency factor is added to the models presenting poor fits ($P < 0.05$; only Konyr_Tobe_300CE). On the top is shown the color legend for the sources tested. (C) Summary of the admixture dates obtained with DATES for the main groups studied. The y axis is the temporal scale from BCE (negative) to CE (positive) dates. The x axis represents the results for the different target groups reported in the legends in each box using the two-way sources reported at the bottom of the three panels formed along the x axis (e.g., source1 + source2). The colored bars represent the date ranges of the culture, while the filled symbols show the admixture dates \pm SEs obtained from DATES and converted into dates considering 29 years per generation starting from the median point of the culture’s age. The three set of sources reported correspond to the summary of the main admixture events described in the text from left to right: the LBA formation of the Scythian gene pools; the BMAC-related influx increasing through time in the Tian Shan Sakas; and the new eastern influx starting in the IA and continuing throughout the centuries. A number-based key (the white numbers from 1 to 6 inside the black circles) connects different tests and analyses shown in the figure with the corresponding arrow in Fig. 4.

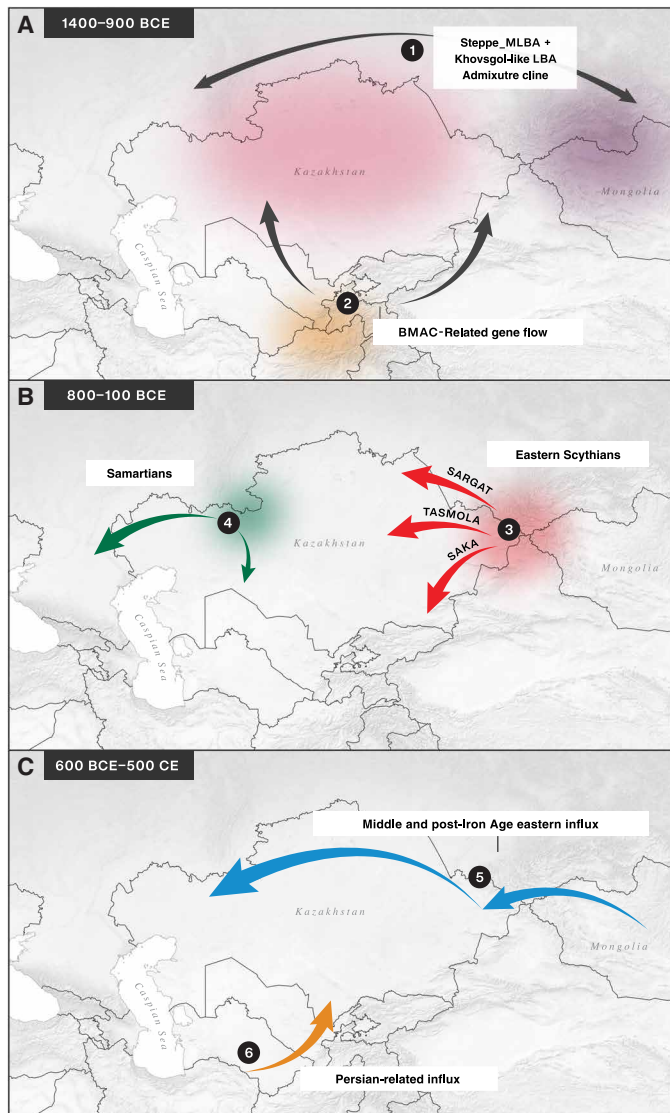


Fig. 4. Summary maps describing the major genetic turnovers that occurred at the turn and throughout the first millennium BCE. (A) Formation of a three-way LBA admixture cline from which (B) eastern Scythian and western Sarmatian gene pools arose and spread throughout the Steppe and (C) a new source of eastern Eurasian ancestry admixing with the Scythian gene pools started in the IA and becoming predominant and widespread at northern latitudes during the Xianbei-Hun period. On the very southern tips of the Steppe, a very different ancestry shift occurred, likely linked with the expansion of the Persian world. The arrows represent the demographic processes analyzed in the present study and are numbered from 1 to 6 to connect them to the main results shown in Fig. 3 from which these inferences have been drawn.

individuals (“Konyr_Tobe_300CE”) fall close to the published Kangju_250CE individuals from a similar time period and region (11), between Sarmatians and BMAC (Fig. 2C). KNT005 is shifted toward BMAC in PCA (Fig. 2C and fig. S1). Furthermore, KNT005 is the only one carrying a South Asian Y haplogroup, L1a2 (data file S1), and showing a South Asian genetic component in ADMIXTURE (Fig. 2D and fig. S2). KNT004 is shifted in PC1 toward East Asians (figs. S1 to S3). Admixture models including ~10% South Asian and ~50% eastern Eurasian influx adequately explain KNT005 and

KNT004, respectively (data file S4). In contrast, the individuals from the site of Alai Nura (Alai_Nura_300CE) in the Tian Shan mountains (~200 km east from the Konyr Tobe site) still lay along the IA cline of the Tian Shan Saka, with four individuals falling closer to Konyr_Tobe_300CE and four closer to the Tasmola/Pazyryk cloud (Fig. 2C and figs. S1 to S3).

Dating ancient admixture

Admixture dating with the DATES program reveal an early formation of the main Scythian gene pools during 1000 to 1500 BCE (Fig. 3C and fig. S4). DATES is designed to model only the two-way admixture, so to account for the estimated three-way models obtained with qpWave and qpAdm, we independently tested the three pairwise comparisons (steppe_MLBA, BMAC, and Khovsgol). DATES was successful in fitting exponential decays for the two western + eastern Eurasian pairs, steppe_MLBA + Khovsgol, and BMAC + Khovsgol, while failing in the western + western Eurasian pair (steppe_MLBA + BMAC) (fig. S4 and table S3). For each target, steppe_MLBA + Khovsgol and BMAC + Khovsgol yielded nearly identical admixture date estimates (table S3). We believe that our estimates mostly reflect an average date between the genetically distinguishable eastern (Khovsgol) and western (steppe_MLBA + BMAC) ancestries, weighted by the relative contribution from the two western sources, rather than reflecting a true simultaneous three-way admixture. It is noteworthy that DATES found increasingly younger admixture dates in the Tian Shan Saka groups as the BMAC-related ancestry increases: from Saka_TianShan_600BCE to the Saka_TianShan_400BCE and especially in the later Alai_Nura_300CE as well as for Pazyryk_Berel_50BCE and Sargat_300BCE with respect to the date of Tasmola_650BCE (~1100 to 900 BCE with respect to ~1300 to 1400 BCE; Fig. 3C). A small-scale gene flow from a BMAC-related source continued over IA may explain both the increase in the BMAC-related ancestry proportion and increasingly younger admixture dates (Fig. 3A). Again, the inferred dates reflect an average over the IA admixture with a BMAC-related source and the LBA one with steppe_MLBA; therefore, they are likely shifted toward older time periods than the actual time of the IA gene flow.

Confirming the results from qpAdm, the admixed individuals from Tasmola_Birlik_640BCE and Korgantas_300BCE (“admixed_Eastern_out_IA”) show very recent admixture dates (Fig. 3C, fig. S4, and table S3). The later groups of Xianbei_Hun_Berel_300CE, Hun_elite_350CE, and Karakaba_830CE further corroborate this trend of recent dates of admixture, revealing that this new eastern influx likely started in the IA and continued at least during the first centuries of the first millennium CE (Fig. 3C, fig. S4, and table S3).

Present-day Kazakhs

PCA, ADMIXTURE, and CHROMOPAINTER/fineSTRUCTURE fine-scale haplotype-based analyses performed on present-day Kazakhs reveal a tight clustering and absence of detectable sub-structure among Kazakhs regardless of the geographic location or Zhuz affiliation (Fig. 2 and fig. S5). We still grouped the Kazakh individuals according to their Zhuz affiliations (which roughly reflects their geographic origin) and ran Globetrotter analyses following the pipeline in (26) as independent replicates to identify the different ancestry sources contributing to the gene pool of Kazakhs and date admixture events. Globetrotter analyses confirmed that the three groups have the same source composition and admixture

dates and are a result of a complex mixture of different western, southern, and eastern Eurasian ancestries (table S4). The dates of admixture identified by Globetrotter highlight a narrow and recent time range for the formation of the present-day Kazakh gene pool, between 1341 and 1544 CE (table S5).

DISCUSSION

Our analysis of more than 100 ancient individuals from Central Asia shows that IA nomad populations of the Kazakh Steppe formed through extensive admixture, resulting from complex interactions between preceding MLBA populations from the steppe and the neighboring regions (Figs. 2A, 3, A and C, and 4A). Our findings shed new light onto the debate about the origins of the Scythian cultures. We do not find support for a western Pontic-Caspian steppe origin, which is, in fact, highly questioned by more recent historical/archeological work (1, 2). The Kazakh Steppe origin hypothesis finds instead a better correspondence with our results, but rather than finding support for one of the two extreme hypotheses, i.e., single origin with population diffusion versus multiple independent origins with only cultural transmission, we found evidence for at least two independent origins as well as population diffusion and admixture (Fig. 4B). In particular, the eastern groups are consistent with descending from a gene pool that formed as a result of a mixture between preceding local steppe_MLBA sources (which could be associated with different cultures such as Sintashta, Srubnaya, and Andronovo that are genetically homogeneous) and a specific eastern Eurasian source that was already present during the LBA in the neighboring northern Mongolia region (27). The genetic structure of the Early IA Tasmola culture of central and northern Kazakhstan is mostly composed of an equal mixture of these two ancestries, although smaller amounts of gene flow from an Iranian-related source are also required (Figs. 3A and 4A and data file S4). We found that overall BMAC-related populations from Turan provide the best fit to our models while Iranian-related sources further to the west, such as the BA groups from the northern Caucasus, fail (data file S4). These results corroborate the historical/archeological hypotheses of a cultural connection between the southern civilizations and the northern steppe people (3). This BMAC influx continues in the later fourth- to first-century BCE-CE Scythian groups from the northeastern Pazyryk site of Berel and becomes increasingly higher and nonuniformly distributed in the southeastern Saka individuals from the Tian Shan mountains (Figs. 2, B and D, 3, A and C, and 4B; fig. S4; table S3; and data file S4).

The two previously published individuals from the Aldy-Bel culture of the Arzhan 2 site in the Tuva region fall within the main eastern Scythian genetic cluster, confirming that it was present also in the same site where the earliest Scythian burials are found (Fig. 2A). These data, coupled with recent findings from the IA transition in Mongolia (28), seem to point to an origin in the Altai area of a main genetic substratum that formed all the eastern Scythians (Fig. 4B). The western Sarmatians from the southern Ural region also formed as a result of admixture between the same three ancestral sources as the eastern Scythians (Fig. 3A). Nevertheless, the eastern Eurasian ancestry is present only in a small amount in Sarmatians (Fig. 3A). In addition, their early admixture dates (Fig. 3C) and the absence of an admixture cline between the Sarmatians and the eastern groups (Fig. 2, A, B, and D) suggest that the Sarmatians descend from a related but different LBA gene pool compared with

the one that contributed to the eastern Scythians (likely differently located along an LBA admixture cline). Given the geographic location of the earliest Sarmatian sites found so far, we hypothesize that this gene pool originated in the LBA southern Ural area (Fig. 4B). More data from the later and westernmost Scythian cultures of the Caucasus and eastern Europe will provide a better understanding of their genetic affinities with the earlier Scythians from the Kazakh Steppe analyzed in the current study. Furthermore, our results show that the northern sedentary Sargat-related cultures show a close genetic proximity with the Scythians especially with the eastern nomad groups (Fig. 2B). The Sargats show additional affinity not found in the Scythian groups ultimately related to a northern Siberian lineage (Figs. 2D and 3A). This is consistent with the historical hypothesis that the Sargat people formed as a result of admixture between incoming Scythian groups and an unsampled local or neighboring population that possibly carried this extra Siberian ancestry (3, 18).

From the second half of the first millennium BCE, we detect a major genetic shift in a number of outliers that are interestingly linked with the emergence of the Korgantas culture that replaced the Tasmola in central Kazakhstan. In particular, we observe an influx from an eastern Eurasian source that is different from the one that contributed to the shift in the LBA (Figs. 3A and 4C and table S2). At the turn of the first millennium CE, this mixed genetic profile became widespread among the northeastern individuals associated with the Xianbei-Hun cultures and the later medieval individuals (Figs. 2, C and D, 3B, and 4C, and table S2). The highly variable admixture proportions and dates obtained for those individuals suggest that this was an ongoing process that characterized the first centuries CE (first to fifth century at least; Fig. 3C, fig. S4, and table S3). Additional genetic data from the first millennium CE will allow a more comprehensive understanding of the nature and the extent of this heterogeneity. Instead, in the southern Kazakhstan region, the individuals from the Konyr Tobe site located in the ancient city of Otyrar Oasis show a different genetic turnover mostly characterized by an increase in Iranian-related genetic ancestry, most likely reflecting the influence of the Persian empires (Fig. 4C) (20, 29). Outliers, with high eastern Eurasian admixture or with gene flow from South Asia, suggest that the population of this city at that time was heterogeneous (Fig. 2C and data file S4). During this period, Otyrar was a main center of the Kangju kingdom and a crossroad along the Silk Road (29). In the neighboring region of the Tian Shan mountains, in the third century CE site of Alai Nura, a genetic profile typical of the much earlier IA Tian Shan Sakas can still be found (Fig. 3B and data file S4).

The heterogeneity and geographic structuring observed during the IA, the Xianbei-Hun, and the medieval periods in Kazakhstan come in strong contrast with the genetic homogeneity observed among present-day Kazakhs (fig. S5). Fine-scale haplotype-based analyses confirmed this homogeneity and showed, in line with previous findings (26), that the Kazakh gene pool is a mixture of different western and eastern Eurasian sources (table S4). Our results on the ancient populations revealed that this was a result of the very complex demographic history, with multiple layers of western and eastern Eurasian ancestries mixing through time. The admixture dates obtained for present-day Kazakhs overlap with the period when the Kazakh Khanate was established (~15th century CE; table S5). Furthermore, the gene pool of present-day Kazakhs cannot be fully modeled as a mixture of post-IA northern Xianbei-Hun and

southern Kangju-related gene pools (data file S4). These findings suggest that recent events, likely enfolding during the second millennium CE, were associated with more demographic turnovers in this region that ultimately lead to the homogenization of the Kazakh gene pool as a consequence of the establishment of the Kazakh Khanate with its strict exogamic rules (21).

MATERIALS AND METHODS

Radiocarbon dating

We selected 48 samples for radiocarbon dating. They were chosen to be representative of the different cultures/genetic clusters observed or key genetic outliers. Additional 14 samples were already ^{14}C -dated in previous studies (text S1), summing up to a total of 62 individuals directly ^{14}C -dated (data file S1). For the new dates, the analyses were done at the Curt-Engelhorn-Zentrum Archaeometry gGmbH, Mannheim, Germany. Collagen was extracted from the bones, purified by ultrafiltration (>30 kDa), and freeze-dried. Then, the samples were combusted to CO_2 in an elemental analyzer, and CO_2 was converted to graphite via catalysis. The $^{14}\text{C}/^{12}\text{C}$ ratio was obtained using a mini radiocarbon dating system–accelerator mass spectrometry. The resulting ^{14}C ages were normalized to $\delta^{13}\text{C} = -25$ per mil (30). The ^{14}C ages were then calibrated (we considered the Cal 2 σ for downstream analyses) using the dataset INTCAL13 (31) and the software SwissCal 1.0 (30).

DNA extraction, library preparations, and sequencing

DNA from the ancient individuals analyzed in this study was obtained following strict sampling and extraction protocols performed in an ancient DNA clean room at the facilities of the Max Planck Institute for the Science of Human History, Jena, Germany. In brief, 40 to 70 mg of bone or tooth powder was used for DNA extraction following a previously published protocol, optimized for the retrieval of short DNA fragments (32). For the initial lysis step, the powder was incubated for 12 to 16 hours (37°C) in 1 ml of extraction buffer containing 0.45 M EDTA (pH 8.0) and proteinase K (0.25 mg/ml) and subsequently purified using a binding buffer containing guanidine hydrochloride, sodium acetate (pH 5.2), and isopropanol (32), in combination with the High Pure Viral Nucleic Acid Large Volume Kit (Roche). Last, DNA extracts were eluted in 100 μl of TET [10 mM tris-HCl, 1 mM EDTA (pH 8.0), and 0.05% Tween 20]. Following DNA extraction, 25 μl of extract from each sample were used to produce double-stranded DNA libraries using a published protocol (33) with an initial treatment using the enzymes uracil-DNA glycosylase (UDG) and endonuclease VIII following a previously described procedure (34). This step allows for the partial removal of uracils resulting from postmortem DNA damage (cytosine deamination), retaining enough damage at the terminal nucleotides of the fragments to permit ancient DNA authentication. The resulting NSG libraries were quantified on a quantitative polymerase chain reaction (qPCR) instrument (LightCycler 96 System, Roche) using the IS7/IS8 primer set and DyNAmo SYBR Green qPCR Kit (Thermo Fisher Scientific) (33). Subsequently, libraries were double-indexed using a combination of indexing primers containing unique 8–base pair (bp) identifiers (35). Ten-cycle indexing PCR reactions were carried out using Pfu Turbo Cx Hotstart DNA Polymerase (Agilent). PCR products were purified using the MinElute DNA purification kit (QIAGEN) and were subsequently qPCR-quantified using the IS5/IS6 primer set (35). Indexed libraries were then amplified

with the IS5/IS6 primer set using the Herculase II Fusion DNA Polymerase (Agilent) to achieve a maximum of 10 copies per reaction, and amplification products were purified using the MinElute DNA purification kit (QIAGEN). Moreover, the concentration (nanograms per microliter) of amplified libraries was measured on an Agilent 4200 TapeStation instrument (Agilent) using the D1000 ScreenTape system (Agilent). Last, an equimolar pool of 69 of 117 UDG-half libraries was prepared for shotgun sequencing within the Max Planck Institute for the Science of Human History facilities on an Illumina HiSeq 4000 platform using a single-end 76-cycle sequencing kit. All the sequenced libraries showed high human endogenous DNA proportions (between 1 and 85% with only one library showing 0.8%) and ancient DNA characteristic damage patterns at the end of the fragments (3' end at least ~0.05%; data file S1). Therefore, all the sample libraries were enriched using DNA probes spanning 1,237,207 genome-wide SNPs known to be variable in human populations. For this, all libraries were reamplified using the Herculase II Fusion DNA Polymerase (Agilent) to achieve 1 to 2 mg of total DNA in 5.2 μl (200 to 400 ng/ μl), they were then purified using the MinElute DNA purification kit (QIAGEN), and their concentrations were measured on a NanoDrop spectrophotometer (Thermo Fisher Scientific). All amplified libraries were subsequently captured following an established in-solution DNA capture protocol (5, 36, 37).

Modern DNA genotyping and quality controls

Genomic DNA from the 96 Kazakh individuals was extracted using the QIAamp DNA Mini Kit (QIAGEN, Germany) according to the manufacturer's protocol. The DNA was quantified spectrophotometrically (Eppendorf BioPhotometer Plus) and fluorometrically (Qubit 2.0). The DNA samples were then genotyped for ~600,000 genome-wide SNPs with the Affymetrix Axiom Genome-wide Human Origins 1 (HO) array platform performed at the ATLAS Biolabs GmbH in Berlin (Germany). Quality controls were performed with PLINK v.1.9 (38). All the 96 individuals had a genotype success rate higher than 95%, and all SNPs had a success rate higher than 95% and were therefore kept for downstream analyses. We then merged our 96 individuals with 18 previously published Kazakh individuals also genotyped on a HumanOrigins array (26). On this dataset ($N = 114$), we estimated recent relatedness values among each pair of individuals with the "--genome" function restricting the analysis only on 73,076 SNPs with low linkage disequilibrium (LD) ($r^2 < 0.1$) that was estimated setting the "--indep-pairwise 50 100 0.1" parameters. We found only two couples with PI-HAT values [i.e., coefficient of relatedness (38)] compatible with a third- to second-degree relation ($0.25 > \text{PI-HAT} > 0.125$) and involved one couple of previously published Kazakh individuals (KZH-1611 and KZH-1750, PI-HAT = 0.23) and one couple formed by a new and a previously published individual (KZH-1650 and E01; PI-HAT = 0.19).

Ancient DNA data processing

Raw data

Demultiplexing of the sequenced reads was done allowing only one mismatch in the indexes. Adaptor removal, mapping to the reference genome, and duplicate removal were done through the EAGER 1.92.32 workflow (39). We used AdapterRemoval v2.2.0 to remove adaptors discarding reads shorter than 30 bp (40). We then mapped the reads to Human Reference Genome Hs37d5 using the bwa v0.7.12 aln/samse alignment algorithm (41) with an edit distance

parameter (“-n”) of 0.01 and a seed length (“-l”) of 32 and keeping only high-quality reads (phred mapping quality of ≥ 30) using Samtools v1.3 (42). We then used DeDup v0.12.2 to remove PCR duplicates (39).

Authentication and contamination estimates

We used mapDamage v2.0 (43) to assess the amount of deamination at the ends of the fragments on a subset of 100,000 high-quality reads using default parameters. We assessed exogenous human DNA contamination levels using ANGSD v0.910 (44) for nuclear (based on X chromosome heterozygosity levels in males) and Schmutzi (45) for mitochondrial DNA contamination. Among the males with enough coverage (i.e., >200 SNPs on X chromosome), none of the individuals had a contamination level of $>7\%$, and only one had $>4\%$. Even if the results are less reliable (i.e., $<<200$ SNPs), we removed from further analyses two males that showed signs of moderate nuclear contamination ($>10\%$) that were also PCA outliers. Furthermore, none of the individuals (males and females) showed levels of mitochondrial contamination of $>3\%$ (data file S1).

Genotyping

We used pileupCaller (<https://github.com/stschiff/sequenceTools>) with the “-randomHaploid” mode to call haploid genotypes for each position captured on the 1240K panel by randomly choosing one high-quality base (phred base quality score of ≥ 30). To call transitions, we first clipped 2 bp from each end of the high-quality reads using the trimBam module of bamUtil v1.0.13 (46) to reduce the numbers of wrong calls due to high deamination at the last two bases, while we used the full high-quality reads to call transversions. At this stage, we excluded from the analyses four individuals with the lowest coverage presenting $<20,000$ SNPs typed on the 1240K panel.

We then merged the newly produced genotype data of 111 ancient individuals with the 96 modern Kazakh and a reference panel composed of 2280 modern individuals genotyped with the Human-Origins array (26, 47, 48) and 959 ancient individuals’ haploid genotypes obtained from a mix of 1240K capture and shotgun sequencing data (4, 6, 7, 9–11, 25–27, 36, 48–52) (data files S2 and S3). This 1240KHO dataset consisting of 586,594 overall SNPs was used for explorative global structure population genetic analyses. For fine-scale ancestry deconvolution and admixture dating analyses, we compiled a higher-coverage 1240K dataset merging only data obtained with the 1240K capture technique or whole-genome shotgun data pooled down to comprehend 1240K sites only (1,233,013 overall SNPs; data files S2 and S3).

Sex determination

Genetic sex was determined calculating the ratio between the coverage on the X and Y chromosomes over the one on the autosomes. We found highly consistent ratios, allowing us to confidently infer the sex of all the individuals. One individual showed a Y/autosomes proportion of 0.96 (data file S1). Since the X-based contamination estimates are extremely low and the X/autosomes ratio within the normal range for males, the most likely explanation is that this individual carries a XYY karyotype. This condition is known as the XYY syndrome and is relatively rare (1 in 1000 births) and largely asymptomatic (53). It commonly affects stature (i.e., increased height) and can slightly influence cognitive or behavioral functions (53).

Genetic relatedness estimation

We assessed relatedness between individuals by calculating the rate of mismatching alleles between every pair of individuals (pairwise mismatch rate) among the overlapping positions as described in

(27, 54). The pairwise mismatch rate provides good evidence of close relationships such as identical individual/twins and first and second degrees (55). We detected a couple of first-degree relatives. The two individuals, a male and a female (ESZ001 and ESZ003), came from the same site (Eleke_Sazy_650BCE) and the same burial (Mound 4). Another first-degree couple was found in Taldy_7cBCE site (TAL003 and TAL004). We then identified a couple of second-degree relatives from the Karashoky_7cBCE site (KSH001 and KSH003) and a couple of possible second- to third-degree relatives between two individuals retrieved from two different Tasmola sites of Akbeit_7cBCE and Nurken_8cBCE (AKB001 and NUR002). We removed one individual per pair of related couples for downstream population-based analyses (data file S1).

Uniparental haplogroup assignment

We used Schmutzi to obtain the consensus sequence of the mitochondrial DNA with a q10 quality cutoff, and we used HaploGrep2 (56) to assign haplogroups. We used yHaplo (57) to assign Y chromosome haplogroups of the male individuals. To obtain Y chromosome genotypes, we used pileupCaller in the “-majorityCall” mode to call the allele supported by most reads for each Y chromosome SNP included in the 1240K panel (data file S1).

Population structure analyses

We applied the smartpca v16000 function in EIGENSOFT v6.0.1 package (58) on the 1240KHO dataset to run PCA with the lsqproject option to project the data of the ancient individuals on top of PCA calculated on the set of modern populations to bypass the high number of missing genotypes in the ancient data that would artificially shift the eigenvectors toward the origin of the axes. We used a set of 150 present-day Eurasian populations on which we projected our newly produced 111 ancient unrelated individuals that passed the quality controls together with other relevant published ancient genomes. We also ran a PCA only on the genotypes of the present-day Kazakh individuals. We then applied ADMIXTURE v1.3.0. (59) unsupervised cluster analyses testing $K = 2$ to $K = 16$ on a set of worldwide ancient and modern individuals. For each K value tested, we performed 10 independent ADMIXTURE runs with a different random seed to check the convergence of log-likelihoods across the different runs. For each K value, we selected for consideration the run with the highest log-likelihood. We also estimated the cross-validation error (CV-err) for each K value to identify the most parsimonious models (i.e., increasing the number of K values does not produce a visible decrease in CV-err) to avoid overfitting. For ADMIXTURE analyses, we removed the variants with a minor allele frequency of <0.01 , and we pruned the dataset, removing all the SNPs in LD with an $r^2 > 0.4$ setting a 200-SNP sliding window with a 25-SNP step using the dedicated commands in PLINK. The pruned dataset consists of 206,728 SNPs, and we further removed from the analysis all the ancient individuals with a missingness rate higher than 95%, which corresponds to including individuals with at least $\sim 10,000$ nonmissing variants on this thinned dataset.

Individual labeling and population grouping criteria

For the new ancient individuals produced in the current study, we first considered a site-based labeling system consisting of site name plus the age expressed in centuries BCE/CE referring to the median of the archeological time range of the site (e.g., Site_Name_400BCE). In the few cases of same sites presenting burials from different and

discontinuous time periods, multiple unique names for the same site exist (e.g., Berel_50BCE and Berel_300CE). Following the recommendations in (60) for grouping the individuals into populations, we used a “mixed system” consisting of the archaeological culture’s name and the archaeological age of the sites included in the group in centuries BCE/CE (e.g., Tasmola_650BCE). To respect the high levels of admixture and genetic variability observed within most of the ancient cultures studied, we identified and excluded as outliers only the individuals that exceeded 2 SD from the median of at least one of the first three PCs within their respective culture group (table S1).

In the cases of cultures represented by one site only, or by few individuals with highly admixed genetic profiles, we retained the site-based labels or used different grouping combinations depending on specific hypotheses and analyses tested as detailed in Results. We extended this labeling system also to previously published ancient individuals belonging to closely related cultures. Nevertheless, to limit potential batch effects due to different laboratory techniques and sequencing methods (i.e., shotgun or 1240K capture), we avoided grouping together our newly generated individuals with previously published ones even if belonging to the same age or culture, opting for using them as independent control groups for validation (data file S3). For consistency with the literature, the rest of the reference ancient and modern population labels and groupings were kept the same as in the original publications unless stated otherwise (data file S3).

F-statistics and ancestry modeling

All the f -statistic–based analyses were run using the dedicated programs in the ADMIXTOOLS package (47) on the 1240K dataset. We run *outgroup-f3* analyses with qp3Pop (v400). We tested the forms $f_3(\text{Test}, X; \text{Mbuti})$ for the Kazakh ancient individuals as *Test* against every other X individual/population included in the dataset. For the newly reported individuals, *outgroup-f3* was run on a site-based grouping considering separate the PCA outliers. To test specific hypotheses detailed in the results, we also computed f_4 -statistics with qpDstat (v711) setting the “f4mode: YES” option. For both f_3 - and f_4 -statistics, we considered only the tests that had a number of overlapping SNPs of $>30,000$, and we considered a $Z > |3|$ as a threshold for significance.

We then run f_4 -statistic–based ancestry decomposition analyses on the 1240K dataset using the qpWave and qpAdm (v632) pipeline (5, 61). SEs for the computed f -statistics were estimated using a block jackknife with a 5-centimorgan block. We used the following set of eight outgroups (OG1) by including representatives of western and eastern Eurasian and relevant non-Eurasian ancient lineages using directly ancient individuals or present-day proxies: Mbuti, Natufian, Anatolia_N, Ganj_Dareh_N, Villabruna, Onge, Ami, and Mixe. As sources, we used, when available, ancient populations from the closest available preceding time periods. We started by selecting proximal source populations for the IA groups by choosing representatives of the three main genetic ancestries found in the MLBA in the Kazakh Steppe and the surrounding regions. Specifically, we used as western ancestry sources a chosen set of steppe_MLBA groups as well as the earlier steppe_EBA (i.e., Yamnaya and Afanasievo) for completeness of analysis. We selected two groups from the west and central clusters described in (7): Sintashta_MLBA and Srubnaya as representative of the western cultures from the southern Ural area (steppe_MLBA_west) and Dali_MLBA and Krasnoyarsk_MLBA from northeastern Kazakhstan and the Minusinsk Basin in Russia,

respectively, as representative of the eastern cluster showing higher affinity to preceding local hunter-gatherer populations and ultimately to the ANE-related ancestry (steppe_MLBA_central). As sources of East Asian ancestries, we used previously published Eneolithic and Early BA individuals from the Baikal region (Baikal_EN and Baikal_EBA) and from the Minusinsk Basin (Okunevo) and LBA individuals from Khovsgol site in northern Mongolia (Khovsgol) as well as Neolithic individuals from the Amur River Basin representative of a deep north East Asian lineage (DevilsCave_N) presenting a high genetic continuity with modern individuals from the same region (52, 62). As third Iranian-related sources, we used the available Eneolithic groups from Iran (Iran_ChL and Hajji_Firuz_C), the MLBA groups from the Caucasus (Caucasus_MBA_North_Caucasus, Caucasus_Late_Maykop, and Caucasus_Kura_Araxes), Armenia_LBA, Eneolithic from Turan (Geoksiur_EN and Tepe_Hissar_C), and BA from Turan associated with the Bactria-Margiana complex or BMAC (Gonur1_BA) and later MLBA “postBMAC” (Sappali_Tepe_BA, Bustan_BA, and Dzarkutan1_BA). We first performed 1648 qpWave/qpAdm-independent tests for each target group, permuting all combinations of two-way ($N = 16$) and three-way ($N = 96$) sources (data file S4). For the IA outliers and the later CE groups, we used the preceding IA groups as first sources and tested different second and eventually third sources (when two-way tests failed) to narrow down closer proxies that could explain the nature of the observed genetic turnovers. For the modeling of present Kazakh Zhuz, we used the 1240KHO dataset using the same set of outgroups used for the ancient (albeit the modern populations are represented by more individuals in the 1240KHO with respect to the 1240K dataset; data file S2). We tested two-way models (data file S4) with all the later CE individuals from the northern latitude, showing the eastern Eurasian influx as the first source (Xiambei_Hun_Berel_300CE, Hun_elite_350CE, Karakaba_830CE, and Kayalyk_950CE) and the southern CE individuals with Iranian-related influx as the second source (Konyr_Tobe_300CE and Alai_Nura_300CE).

Admixture dating

We used DATES (7) on the 1240K dataset to date the admixture events identified from previous analyses in the IA and post-IA Kazakh individuals. The method is conceptually similar to commonly used admixture-dating methods based on LD such as ALDER (63), although instead of calculating the weighted LD decay, which would require high coverage with virtually no missing data, DATES use the decay of ancestry covariance (AC) coefficients between pairs of overlapping SNPs over increasing genetic distance. As for the LD-based methods, an exponential function can be fitted to the decay of weighted AC as genetic distance increases to infer admixture parameters such as the number of generations since admixture (63). We then considered a standard 29 years per generation to convert the generation times into years since admixture (7). DATES assumes a two-way admixture; therefore, we used as sources pairwise combinations of the best proxies resulted from qpAdm modeling analyses. In choosing the source populations, we also considered that the method is sensitive to sample sizes and coverage, preferring proxies with a higher number of individuals or, when possible, pooling together genetically homogeneous populations to reduce statistical noise (table S3). In choosing the target individuals, for sites with more complex chronology (i.e., containing burials belonging to different time periods), we included only individuals directly ^{14}C -dated to reduce the errors due to incorrect context dating.

CHROMOPAINTER and fineSTRUCTURE analyses

We reconstructed the phase of haplotypes for the modern Kazakh individuals together with the set of worldwide modern populations present in the 1240KHO dataset using SHAPEIT2 v2.r790 (64) with default parameters and using HapMap phase 3 recombination maps. To explore the fine-scale population structure among present-day Kazakh individuals, we applied the haplotype-based CHROMOPAINTERv2/fineSTRUCTURE pipeline (65) on the Central Asian Southern Steppe populations present in the dataset plus all the Kazakh individuals (96 new and 18 previously published). We first estimated the mutation/emission and the switch rate parameters with 10 steps of the Expectation-Maximization algorithm on a subset of chromosomes {4, 10, 15, 22} using each individual as “donor” and “recipient.” Then, we averaged the values across chromosomes (normalized by the number of SNPs per chromosome) and individuals, and we used these mutation/emission and switch rate parameters to run CHROMOPAINTER again on all chromosomes, considering a parameter $k = 50$ to specify the number of expected chunks to define a region. The obtained matrix of haplotype-sharing “chunk” counts was summed up across all the 22 autosomes and submitted to the fineSTRUCTURE clustering algorithm version fs2.1 (65). We ran fineSTRUCTURE pipeline by setting 1,000,000 “burn-in” Markov chain Monte Carlo iterations, followed by additional 2,000,000 iterations and sampling the inferred clustering patterns every 10,000 runs. Last, we set 1,000,000 additional hill-climbing steps to improve posterior probability and merge clusters in a stepwise fashion. Individuals were hierarchically assembled into clusters until reaching the final configuration tree. We then applied the GLOBETROTTER algorithm (66) using the Kazakhs, grouped according to their Zhuz affiliation, as targets and a set of 85 non-Inner Eurasian populations as reference groups following the same pipeline detailed in (26) to date admixture and identify the main contributing sources. All GLOBETROTTER runs were conducted according to the guidelines reported in (66) and performing a first run standardizing over a null individual.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/13/eabe4414/DC1>

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

REFERENCES AND NOTES

- L. Koryakova, Europe to Asia, in *The Oxford Handbook of the European Iron Age*, C. Haselgrove, K. Rebay-Salisbury, P. S. Wells, Eds. (Oxford Univ. Press, 2018), pp. 1–41.
- J. Davis-Kimball, L. Koryakova, E. M. Murphy, L. T. Yablonsky, *Kurgans, Ritual Sites, and Settlements: Eurasian Bronze and Iron Age* (British Archaeological Reports Limited, 2000), 324 pp.
- L. Koryakova, A. Epimakhov, The Urals and Western Siberia in the Bronze and Iron Ages. 10.1017/cbo9780511618451 (2007).
- M. E. Allentoft, M. Sikora, K.-G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. D. Casa, P. Dąbrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtychyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, Ł. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Soenov, V. Szevevényi, G. Tóth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepikoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
- W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfalk, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgrén, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. R. Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
- P. de Barros Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaubert, V. Merz, A. Biddanda, I. Merz, V. Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediy, I. Ullah, K.-G. Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A. Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. McKenzie, A. Margaryan, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E. Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
- V. M. Narasimhan, N. Patterson, P. Moorjani, N. Rohland, R. Bernardos, S. Mallick, I. Lazaridis, N. Nakatsuka, I. Olalde, M. Lipson, A. M. Kim, L. M. Olivieri, A. Coppa, M. Vidale, J. Mallory, V. Moiseyev, E. Kitov, J. Monge, N. Adamski, N. Alex, N. Broomandkoshbacht, F. Candilio, K. Callan, O. Cheronet, B. J. Culleton, M. Ferry, D. Fernandes, S. Freilich, B. Gamarra, D. Gaudio, M. Hajdinjak, É. Harney, T. K. Harper, D. Keating, A. M. Lawson, M. Mah, K. Mandl, M. Michel, M. Novak, J. Oppenheimer, N. Rai, K. Sirak, V. Slon, K. Stewardson, F. Zalza, Z. Zhang, G. Akhatov, A. N. Bagashev, A. Bagnera, B. Baitanayev, J. Bendezu-Sarmiento, A. A. Bissembaev, G. L. Bonora, T. T. Charyginov, T. Chikisheva, P. K. Dashkovskiy, A. Derevianko, M. Dobeš, K. Douka, N. Dubova, M. N. Duisengali, D. Enshin, A. Epimakhov, A. V. Fribus, D. Fuller, A. Goryachev, A. Gromov, S. P. Grushin, B. Hanks, M. Judd, E. Kazizov, A. P. Krygin, E. Kupriyanova, P. Kuznetsov, D. Luiselli, F. Maksudov, A. M. Mamedov, T. B. Mamirov, C. Meiklejohn, D. C. Merrett, R. Micheli, O. Mochalov, S. Mustafokulov, A. Nayak, D. Pettener, R. Potts, D. Razhev, M. Rykun, S. Sarno, T. M. Savenkova, K. Sikhymbaeva, S. M. Slepchenko, O. A. Soltobaev, N. Stepanova, S. Svyatko, K. Tabaldiev, M. Teschler-Nicola, A. A. Tishkin, V. V. Tkachev, S. Vasilyev, P. Velemínský, D. Voyakin, A. Yeremolayeva, M. Zahir, V. S. Zubkov, A. Zubova, V. S. Shinde, C. Lalueza-Fox, M. Meyer, D. Anthony, N. Boivin, K. Thangaraj, D. J. Kennett, M. Frachetti, R. Pinhasi, D. Reich, The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
- B. van Geel, N. A. Bokovenko, N. D. Burova, K. V. Chugunov, V. A. Dergachev, V. G. Dirksen, M. Kulkova, A. Nagler, H. Parzinger, J. van der Plicht, S. S. Vasiliev, G. I. Zaitseva, The sun, climate change and the expansion of the Scythian culture after 850 BC. *NATO Sci. Ser.* **42**, 151–158 (2004).
- M. Unterländer, F. Palstra, I. Lazaridis, A. Piliipenko, Z. Hofmanová, M. Groß, C. Sell, J. Blöcher, K. Kirsanov, N. Rohland, B. Rieger, E. Kaiser, W. Schier, D. Pozdniakov, A. Khokhlov, M. Georges, S. Wilde, A. Powell, E. Heyer, M. Currat, D. Reich, Z. Samashev, H. Parzinger, V. I. Molodin, J. Burger, Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat. Commun.* **8**, 14615 (2017).
- M. Krzewińska, G. M. Kilinc, A. Juras, D. Koptekin, M. Chyleńska, A. G. Nikitin, N. Shcherbakov, I. Shuteleva, T. Leonova, L. Kraeva, F. A. Sungatov, A. N. Sultanova, I. Potekhina, S. Łukasik, M. Krenz-Niedbala, L. Dalén, V. Sinika, M. Jakobsson, J. Stora, A. Götherström, Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci. Adv.* **4**, eaat4457 (2018).
- P. de B. Damgaard, N. Marchi, S. Rasmussen, M. Peyrot, G. Renaud, T. Korneliusen, J. V. Moreno-Mayar, M. W. Pedersen, A. Goldberg, E. Usmanova, N. Baimukhanov, V. Loman, L. Hedeager, A. G. Pedersen, K. Nielsen, G. Afanasiev, K. Akmatov, A. Aldashev, A. Alpaslan, G. Baimbetov, V. I. Bazaliiskii, A. Beisenov, B. Boldbaatar, B. Boldgiv, C. Dorzhu, S. Ellingvag, D. Erdenebaatar, R. Dajani, E. Dmitriev, V. Evdokimov, K. M. Frei, A. Gromov, A. Goryachev, H. Hakonarson, T. Hegay, Z. Khachatryan, R. Khaskhanov, E. Kitov, A. Kolbina, T. Kubatbek, A. Kukushkin, I. Kukushkin, N. Lau, A. Margaryan, I. Merkyte, I. V. Mertz, V. K. Mertz, E. Mijiddorj, V. Moiseyev, G. Mukhtarova, B. Nurmukhanbetov, Z. Orozbekova, I. Panyushkina, K. Pieta, V. Smrčka, I. Shevina, A. Logvin, K.-G. Sjögren, T. Štolcová, A. M. Taravella, K. Tashbaeva, A. Tkachev, T. Tulegenov, D. Voyakin, L. Yepikoposyan, S. Undrakhbold, V. Varfolomeev, A. Weber, M. A. W. Sayres, N. Kradin, M. E. Allentoft, L. Orlando, R. Nielsen, M. Sikora, E. Heyer, K. Kristiansen, E. Willerslev, 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
- M. Järve, L. Saag, C. L. Scheib, A. K. Pathak, F. Montinaro, L. Paganí, R. Flores, M. Guellil, L. Saag, K. Tambets, A. Kushniarevich, A. Solnik, L. Varul, S. Zadnikov, O. Petrauskas, M. Avramenko, B. Magomedov, S. Didenko, G. Toshev, I. Bruyalko, D. Grechko, V. Okatenko, K. Gorbenko, O. Smyrnov, A. Heiko, R. Reid, S. Sapielin, S. Sirotnin, A. Tairou, A. Beisenov, M. Starodubtsev, V. Vasiliev, A. Nechvaloda, B. Atabiev, S. Litvinov, N. Ekomasova, M. Dzhaubermezov, S. Voroniatov, O. Utevska, I. Shramko, E. Khusnutdinova, M. Metspalu, N. Savelev, A. Kriiska, T. Kivisild, R. Villemes, Shifts in the genetic landscape of the Western Eurasian Steppe Associated with the beginning and end of the Scythian Dominance. *Curr. Biol.* **29**, 2430–2441.e10 (2019).
- A. Y. Alekseev, A. Y. Alekseev, N. A. Bokovenko, Y. Boltrik, K. A. Chugunov, G. Cook, V. A. Dergachev, N. Kovalyukh, G. Possnert, J. van der Plicht, E. M. Scott, A. Semestsov, V. Skripkin, S. Vasiliev, G. Zaitseva, A chronology of the scythian antiquities of Eurasia based on new archaeological and 14C data. *Radiocarbon* **43**, 1085–1107 (2001).

14. V. Mordvinceva, S. Reinhold, The northern Black Sea and North Caucasus, in *The Oxford Handbook of the European Iron Age*, C. Haselgrove, K. Rebay-Salisbury, P. S. Wells, Eds. (Oxford Univ. Press, 2018), pp. 1–50.
15. Deutsches Archäologisches Institut, Museum für Vor- und Frühgeschichte (Berlin, Germany), G. K. Hypo-Kulturstiftung (Munich), G. Martin-Gropius-Bau, Berlin), *Im Zeichen des goldenen Greifen: Königsgräber der Skythen* (Prestel, Munich, 2007).
16. A. Farkas, Frozen tombs of Siberia: The Pazyryk burials of Iron Age horsemen. By Sergei I. Rudenko. Translated, with a preface, by M. W. Thompson. Berkeley and Los Angeles: University of California Press, 1970. xxxvi, 340 pp. 33 color, 147 black and white plates. \$30.00. *Slav. Rev.* **30**, 657–658 (1971).
17. M. E. Hall, Towards an absolute chronology for the Iron Age of Inner Asia. *Antiquity* **71**, 863–874 (1997).
18. A. M. K. Lerner, *Iron Age Nomads of the Urals: Interpreting Sauro-Sarmatian and Sargat Identities* (JMI ProQuest, Ann Arbor, 2006).
19. The Routledge Handbook of Archaeology and Globalization (2016); 10.4324/9781315449005.
20. J. Bemmann, Complexity of interaction along the Eurasian Steppe Zone in the first millennium CE. *Bonn Contrib. Asian Archeol.* **7**, 708 (2015).
21. I. Bashiri, Martha Brill Olcott. The Kazaks. Second Edition. Stanford, CA: Hoover Institution Press, 1995. xxiii, 388 pp. *Can. Am. Slavic Stud.* **31**, 237–239 (1997).
22. S. M. Akimbekov, *History of the Steppes: The Phenomenon of the State of Genghis Khan in the History of Eurasia* (Center of Asia, 2011).
23. S. Kenzheakmetuli, in *Seven Treasures: Collection of 2 Books*, A. A. Tili, Ed. (2002), pp. 134–135.
24. Ä. Z. Beisenov, D. Duisenbay, I. Akhiyarov, G. Sargizova, Dromos Burials of Tasmola Culture in Central Kazakhstan. *Anthropologist* **26**, 25–33 (2017).
25. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr., L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
26. C. Jeong, O. Balanovska, E. Lukianova, N. Khabatkyzy, P. Flegontov, V. Zaporozhchenko, A. Immel, C.-C. Wang, O. Ixan, E. Khussainova, B. Bekmanov, V. Zaibert, M. Lavryashina, E. Pocheshkhova, Y. Yusupov, A. Agdzhozoy, S. Koshel, A. Bukin, P. Nymadawa, S. Turdikulova, D. Dalimova, M. Churnosov, R. Skhalyakho, D. Daragan, Y. Bogunov, A. Bogunova, A. Shtrunov, N. Dubova, M. Zhabagin, L. Yepiskoposyan, V. Churakov, N. Pislegin, L. Damba, L. Saroyants, K. Dibirova, L. Atramantova, O. Utevska, E. Idrisov, E. Kamenshchikova, I. Evseeva, M. Metspalu, A. K. Outram, M. Robbeets, L. Djansugurova, E. Balanovska, S. Schiffels, W. Haak, D. Reich, J. Krause, The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* **3**, 966–976 (2019).
27. C. Jeong, S. Wilkin, T. Amgalantugs, A. S. Bouwman, W. T. T. Taylor, R. W. Hagan, S. Bromage, S. Tsolmon, C. Trachsel, J. Grossmann, J. Littleton, C. A. Makarewicz, J. Krigbaum, M. Burri, A. Scott, G. Davaasambuu, J. Wright, F. Irmer, E. Myagmar, N. Boivin, M. Robbeets, F. J. Rühli, J. Krause, B. Frohlich, J. Hendy, C. Warinner, Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11248–E11255 (2018).
28. C. Jeong, K. Wang, S. Wilkin, W. T. T. Taylor, B. K. Miller, S. Ulziibayar, R. Stahl, C. Chiovelli, J. H. Bemmann, F. Knolle, N. Krادين, B. A. Bazarov, D. A. Miyagashev, P. B. Kononov, E. Zhambaltarova, A. V. Miller, W. Haak, S. Schiffels, J. Krause, N. Boivin, E. Myagmar, J. Hendy, C. Warinner, A dynamic 6,000-year genetic history of Eurasia's Eastern Steppe. *bioRxiv* 2020.03.25.008078, (2020).
29. T. R. Hermes, M. D. Frachetti, E. A. Bullion, F. Maksudov, S. Mustafokulov, C. A. Makarewicz, Urban and nomadic isotopic niches reveal dietary connectivities along Central Asia's Silk Roads. *Sci. Rep.* **8**, 5177 (2018).
30. M. Christl, C. Vockenhuber, P. W. Kubik, L. Wacker, J. Lachner, V. Alfimov, H.-A. Synal, The ETH Zurich AMS facilities: Performance parameters and reference materials. *Nucl. Instrum. Methods Phys. Res., Sect. B* **294**, 29–38 (2013).
31. P. J. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. B. Ramsey, C. E. Buck, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Hafidason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, G. F. Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
32. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
33. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* **2010**, db.prot5448 (2010).
34. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B* **370**, 20130624 (2015).
35. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3–e3 (2012).
36. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
37. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013).
38. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
39. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Knipf, J. Krause, K. Nieselt, EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
40. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC. Res. Notes* **9**, 88 (2016).
41. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
42. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
43. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
44. T. S. Korneliusson, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
45. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
46. G. Jun, M. K. Wing, G. R. Abecasis, H. M. Kang, An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).
47. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
48. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Forbes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Shved, A.-S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoun-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyann, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
49. H. McColl, F. Racimo, L. Vinner, F. Demeter, T. Gakuhari, J. V. Moreno-Mayar, G. van Driem, U. G. Wilken, A. Sequin-Orlando, C. de la Fuente Castro, S. Wasef, R. Shoocongdej, V. Soukavaty, T. Sayavongkhamdy, M. M. Saidin, M. E. Allentoft, T. Sato, A.-S. Malaspinas, F. A. Aghakhanian, T. Korneliusson, A. Prohaska, A. Margaryan, P. de Barros Damgaard, S. Kaewstuthi, P. Lertrit, T. M. H. Nguyen, H.-C. Hung, T. M. Tran, H. Nghia Truong, G. H. Nguyen, S. Shahidan, K. Wiradnyana, H. Matsumae, N. Shigehara, M. Yoneda, H. Ishida, T. Masuyama, Y. Yamada, A. Tajima, H. Shibata, A. Toyoda, T. Hanihara, S. Nakagome, T. Deviese, A.-M. Bacon, P. Düringer, J.-L. Ponche, L. Shackelford, E. Patole-Edoumba, A. T. Nguyen, B. Bellina-Pryce, J.-C. Galipaud, R. Kinaston, H. Buckley, C. Pottier, S. Rasmussen, T. Higham, R. A. Foley, M. M. Lahr, L. Orlando, M. Sikora, M. E. Phipps, H. Oota, C. Higham, D. M. Lambert, E. Willerslev, The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018).
50. C.-C. Wang, S. Reinhold, A. Kalmykov, A. Wissgott, G. Brandt, C. Jeong, O. Cheronet, M. Ferry, E. Harney, D. Keating, S. Mallick, N. Rohland, K. Stewardson, A. R. Kantorovich, V. E. Maslov, V. G. Petrenko, V. R. Erlikh, B. C. Atabiev, R. G. Magomedov, P. L. Kohl, K. W. Alt, S. L. Pichler, C. Gerling, H. Meller, B. Vardanyan, L. Yeganyan, A. D. Rezepkin, D. Mariaschk, N. Berezina, J. Gresky, K. Fuchs, C. Knipper, S. Schiffels, E. Balanovska, O. Balanovska, I. Mathieson, T. Higham, Y. B. Berezin, A. Buzhilo, V. Trifonov, R. Pinhasi, A. B. Belinskij, D. Reich, S. Hansen, J. Krause, W. Haak, Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
51. T. C. Lammidis, K. Majander, C. Jeong, E. Salmela, A. Wessman, V. Moiseyev, V. Khartanovich, O. Balanovsky, M. Ongyther, A. Weihmann, A. Sajantila, J. Kelso, S. Pääbo,

- P. Onkamo, W. Haak, J. Krause, S. Schiffels, Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* **9**, 5018 (2018).
52. M. Sikora, V. V. Pitulko, V. C. Sousa, M. E. Allentoft, L. Vinner, S. Rasmussen, A. Margaryan, P. de Barros Damgaard, C. de la Fuente, G. Renaud, M. A. Yang, Q. Fu, I. Dupanloup, K. Giampoudakis, D. Nogués-Bravo, C. Rahbek, G. Kroonen, M. Peyrot, H. McColl, S. V. Vasilyev, E. Veselovskaya, M. Gerasimova, E. Y. Pavlova, V. G. Chasnyk, P. A. Nikolskiy, A. V. Gromov, V. I. Khartanovich, V. Moiseyev, P. S. Grebenyuk, A. Y. Fedorchenko, A. I. Lebedintsev, S. B. Slobodin, B. A. Malyarchuk, R. Martiniano, M. Meldgaard, L. Arppe, J. U. Palo, T. Sundell, K. Mannermaa, M. Putkonen, V. Alexandersen, C. Primeau, N. Baimukhanov, R. S. Malhi, K.-G. Sjögren, K. Kristiansen, A. Wessman, A. Sajantila, M. M. Lahr, R. Durbin, R. Nielsen, D. J. Meltzer, L. Excoffier, E. Willerslev, The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188 (2019).
 53. I. W. Kim, A. C. Khadilkar, E. Y. Ko, E. S. Sabanegh Jr., 47,XYX Syndrome and Male Infertility. *Rev. Urol.* **15**, 188–196 (2013).
 54. D. J. Kennett, S. Plog, R. J. George, B. J. Culleton, A. S. Watson, P. Skoglund, N. Rohland, S. Mallick, K. Stewardson, L. Kistler, S. A. LeBlanc, P. M. Whiteley, D. Reich, G. H. Perry, Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
 55. J. M. Monroy Kuhn, M. Jakobsson, T. Günther, Estimating genetic kin relationships in prehistoric populations. *PLOS ONE* **13**, e0195491 (2018).
 56. A. Kloss-Brandstätter, D. Pachter, S. Schönheir, H. Weissensteiner, R. Binna, G. Specht, F. Kronenberg, HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).
 57. G. D. Poznik, Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv*, 088716 (2016).
 58. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
 59. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
 60. S. Eisenmann, E. Bánffy, P. van Dommelen, K. P. Hofmann, J. Maran, I. Lazaridis, A. Mittnik, M. McCormick, J. Krause, D. Reich, P. W. Stockhammer, Reconciling material cultures in archaeology with genetic data: The nomenclature of clusters emerging from archaeogenomic analysis. *Sci. Rep.* **8**, 13003 (2018).
 61. D. Reich, N. Patterson, D. Campbell, A. Tandon, S. Mazieres, N. Ray, M. V. Parra, W. Rojas, C. Duque, N. Mesa, L. F. García, O. Triana, S. Blair, A. Maestre, J. C. Dib, C. M. Bravi, G. Bailliet, D. Corach, T. Hünemeier, M. C. Bortolini, F. M. Salzano, M. L. Petzl-Erler, V. Acuña-Alonso, C. Aguilar-Salinas, S. Canizales-Quintero, T. Tusié-Luna, L. Ribba, M. Rodríguez-Cruz, M. Lopez-Alarcón, R. Coral-Vazquez, T. Canto-Cetina, I. Silva-Zolezzi, J. C. Fernandez-Lopez, A. V. Contreras, G. Jimenez-Sanchez, M. J. Gómez-Vázquez, J. Molina, A. Carracedo, A. Salas, C. Gallo, G. Poletti, D. B. Witonsky, G. Alkorta-Aranburu, R. I. Sukernik, L. Osipova, S. A. Fedorova, R. Vasquez, M. Villena, C. Moreau, R. Barrantes, D. Pauls, L. Excoffier, G. Bedoya, F. Rothhammer, J.-M. Dugoujon, G. Larrouy, W. Klitz, D. Labuda, J. Kidd, K. Kidd, A. Di Rienzo, N. B. Freimer, A. L. Price, A. Ruiz-Linares, Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
 62. V. Siska, E. R. Jones, S. Jeon, Y. Bhak, H.-M. Kim, Y. S. Cho, H. Kim, K. Lee, E. Veselovskaya, T. Balueva, M. Gallego-Llorente, M. Hofreiter, D. G. Bradley, A. Eriksson, R. Pinhasi, J. Bhak, A. Manica, Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* **3**, e1601877 (2017).
 63. P.-R. Loh, M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, B. Berger, Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
 64. O. Delaneau, B. Howie, A. J. Cox, J.-F. Zagury, J. Marchini, Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
 65. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012).
 66. G. Hellenthal, G. B. J. Busby, G. Band, J. F. Wilson, C. Capelli, D. Falush, S. Myers, A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
 67. O. Bottyán, Data to the anthropology of the Hun Period population in Hungary. *Annales historico-naturales Musei nationalis hungarici* **59**, 455–464 (1967).
 68. M. Nagy, A Hun-Age burial with male skeleton and horse bones found in Budapest, in *Neglected Barbarians*, F. Curta, Ed. (Studies in the Early Middle Ages, Brepols Publishers, 2010), vol. 32, pp. 137–175.
 69. A. Z. Beisenov, Burial and ritual complex “Kurgan 37 Warriors,” in *Bulletin of SUSU* (Series “Social and Human Sciences,” South Ural State University, Chelyabinsk, 2015) (in Russian), vol. 15, pp. 6–12.
 70. A. Z. Beisenov, E. P. Kitov, Posthumous Skull Trepanation in Elite Burial Grounds of the Saka Period in Central Kazakhstan (FSBEIHE, Altai State University, 2014) (in Russian); 10.14258/izvasu.
 71. A. Z. Beisenov, Earrings of the Saka epoch. *Bulletin of Tomsk State University. History* **6** (2014) (in Russian), pp. 121–128.
 72. E. P. Kitov, A. Z. Beisenov, Skulls with trepanations from barrows of the early iron age of Saryarka. *Bulletin of Moscow University. Series XXIII. Anthropol. Forum* **1**, 37–48 (2015).
 73. A. Z. Beisenov, Tasmola culture of Central Kazakhstan in the studies of the beginning of the XXI century. *Archeol. History Ukraine* **2**, 386–396 (2018).
 74. A. Z. Beisenov, R. S. Minasyan, Mirror from the Tasmola burial ground Birlík. *Bull. Dagestan Sci. Center* **58**, 58–62 (2015).
 75. A. Z. Beisenov, Tasmola burials of the Kyzyl burial ground. Modern solutions of actual problems of Eurasian archeology. *Collect. Sci. Articles* **2**, 139–149 (2018).
 76. A. Z. Beisenov, A. E. Kasenalin, Kyzylshilik burial ground of the Tasmola culture. *Foreign Archeol.* **4**, (2018).
 77. A. Z. Beisenov, Dromos kurgans of the Saka time on the Zharly river (central Kazakhstan). *Samara Scientific Herald* **3**, 77–86 (2016).
 78. A. Z. Beisenov, A. O. Ismagulova, E. P. Kitov, A. O. Kitova, *The Population of Central Kazakhstan in the 1st Millennium BC* (Almaty Institute of Archeology Named after A.Kh. Margulan, 2015) (in Russian), p. 188.
 79. A. Z. Beisenov, E. P. Kitov, The Taldy II burial ground of Tasmola Culture in the Central Kazakhstan (craniological analysis). *Science Journal of Volgograd State University History. Area Studies. International Relations*, **4** 71–85 (2014).
 80. Z. Samashev, E. M. Kariev, S. E. Erbolatov, Hun-Syambian cultural and chronological horizon of Berel, in *Materials of the International Archaeological Conferences*, B.A. Baytanayev, Ed. (L.N. Gumilyov Eurasian National University, 2019) (in Russian), pp. 385–394.
 81. V. E. Stoyanov, A. G. Degtyarev, Report on the Kurgansk detachment of the Ural university expedition of 1963, *Archive of IA RAS* (Institute of Archaeology of the Russian Academy of Sciences), P-1, No. 2749 (in Russian).
 82. Z. S. Samashev, G. S. Zhumabekova, A. S. Ermolaeva, G. Omarov, Military archeology. Weapon and military affair in a historical and social perspective. Early Saka arrowheads from Kazakhstan Altai - SPb (State Hermitage, 1998) (in Russian), pp. 155–160.
 83. Z. S. Samashev, A. S. Ermolaeva, G. S. Zhumabekova, Kazakh Altai in the 1st millennium BC. Kazakhstan in the Saka era, in *Collective Monograph. Almaty* (Institute of Archeology Named after A.Kh. Margulan, 2017) (in Russian), pp. 101–156.
 84. A. Z. Beisenov, V. V. Varfolomeev, V. K. Merz, I. V. Merz, Excavations of the Karaoba burial ground in 2013 (preliminary report). The dialogue of cultures of Eurasia in the archeology of Kazakhstan, in *A Collection of Scientific Articles Dedicated to the 90th Birthday of the Outstanding Archaeologist K. A. Akishev*, T.S. Sadykov, Ed. (Saryarka Publishing House, 2014) (in Russian), p. 736.
 85. A. Onggaruly, G. Kiyasbek, M. Kyzirkhanov, A. Kairmagambetov, The sanctuary of Aigyry 2 in Mangystau (preliminary results), in *Religion and the Worldview System of Ancient and Medieval Nomads of Eurasia. Collection of scientific articles*, A. Onggaruly, Ed. (Institute of Archeology named after A.Kh. Margulan, 2016) (in Russian), pp. 93–106.
 86. A. Onggaruly, The image of the sarmatian batyr by the materials of the sacred Aral-Caspian streams. Heritage of the Great Steppe: masterpieces of jewelry art. **IV. The world of art images of nomads. The exhibition catalogue** (The National Museum of the Republic of Kazakhstan, 2018), pp. 47–51.
 87. A. Onggaruly, V. S. Olkhovskiy, A. Astafyev, D. Darmenov, *The Ancient Sanctuaries of Ustyurt and Eastern Aral* (Institute of Archaeology Named after A.Kh. Margulan, 2017) (in Russian), p. 320.
 88. N. P. Petrov, V. V. Rodionov, Report on archaeological exploration in the Aktobe region in the summer of 1977, in *Funds of AOIKM. Aktyubinsk* (Aktobe Regional Museum of Local History, 1978).
 89. S. Yu Gutsalov, G. V. Makarevich, Report on archaeological explorations and excavations in the Aktobe region in the summer and fall of 1986, in *Funds of AOIKM. Aktyubinsk* (Aktobe Regional Museum of Local History, 1987).
 90. M. G. Moshkova, G. V. Kushaev, Report on the work of the West Kazakhstan expedition of 1969. *Archive of IA RAS* (Institute of Archaeology of the Russian Academy of Sciences), P-1, Nos. 4381 and 4381a (in Russian).
 91. A. A. Bisembaev, A. I. Havansky, Excavations of the burial ground Kajynbulak II in Aktyubinsk region of the Republic of Kazakhstan in 2017. *International scientific conference “New in the research of the early Iron Age of Eurasia: problems, discoveries, methods”*: abstracts. Ans. ed. A.A. Malyshev (MAX Press, 2018) (in Russian), p. 174.
 92. A. M. Mamedov, E. P. Kitov, Funeral ceremony of early nomads from the top Ilek (on the materials of the burial ground Sapibulak). *News Natl. Acad. Sci. Republic Kazakhstan* **6**, 19–60 (2015).
 93. A. Z. Beisenov, L. Kreshioli, G. S. Jumabekova, G. A. Bazarbayeva, E. Barinova, The Early Iron age Burial Ground Kaspan-6 in Jetysay. *Theor. Pract. Archaeol. Res.* **2**, 97–108 (2017).
 94. A. Z. Beisenov, D. B. Duisenbay, E. P. Kitov, M. A. Kulkova, Study of the Saka time barrows in the Kaspan valley in the Zhetyysu. *Foreign Archeol.* **3**, 10 (2018).
 95. A. A. Bisembaev, A. I. Khavansky, M. N. Duisengali, A. M. Mamedov, N. M. Bairov, V. A. Amelin, Research of monuments of the Hun time in the Aktobe region in 2018 (preliminary report). *Archaeol. Kazakhstan.* **1-2**, 235–244 (2018).
 96. Z. Samashev, Monuments of the medieval nomads of Karakaba river in Kazakh Altai. *Archeol. Middle Ages*, 379–409 (2016).
 97. K. M. Baipakov, J. K. Taimagambetov, *Archeology of Kazakhstan: Textbook FOR University Students* (Kazakh University, 2006) (in Russian), p. 355.

98. A. Saipov, S. S. Akylbek, A. K. Avizova, A. S. Maisupova, Archaeological study of necropolis of the industrial oasis. *Izvestiya of the Samara Science Centre of the Russian Academy of Sciences. Social, humanitarian, medicobiological sciences* **21**, 77–85 (2019).
99. B. Ayagan, Kazakhstan. National encyclopedia. 1 (LLP, Kazakh Encyclopedia, 2004) (in Russian), p. 560.
100. E. Maanaev, V. Ploskikh, *On the Roof of the World: Historical Essays on the Pamir-Alai Kirghiz* (Mektep, 1983) (in Russian), p. 144.
101. M. V. Voevodovsky, M. P. Gryaznov, U-Sun burial grounds on the territory of the Kyrgyz SSR. *Bull. Ancient History* **3**, 162–179 (1938).
102. B. A. Ribakov, Steppe zone of the Asian part of the USSR in Scythian-Sarmatian time. *Ser. Archeol. USSR* **10**, 494 (1992).
103. V. A. Mogilnikov, Report on the work of the Irtysh detachment of the West Siberian expedition of 1967. *Archive of IA RAS* (Institute of Archaeology of the Russian Academy of Sciences), P-1, No. 3464 (in Russian).
104. V. A. Mogilnikov, Report on the work of the Irtysh detachment of the West Siberian expedition of 1968. *Archive of IARAS* (Institute of Archaeology of the Russian Academy of Sciences), P-1, No. 3716 (in Russian).
105. V. A. Mogilnikov, To the question of Sargat culture, in *Problems of Archeology and the Ancient History of the Ugrians*, A.P. Smirnov, V.N. Chernetsov, I.F. Erdeli, Eds. (Nauka, 1972) (in Russian), pp. 66–87.
106. V. F. Gening, Ural archaeological expedition of 1961. *Archive of IARAS* (Institute of Archaeology of the Russian Academy of Sciences), P-1, No. 2362 (in Russian).
107. V. A. Buldashev, Funeral rites of the Gorokhov culture. *Abstract. Cand. Diss. Novosibirsk* (Ph.D. thesis, Institute of History and Archaeology, Ural Branch, Russian Academy of Sciences, 1998) (in Russian).
108. V. E. Stoyanov, V. N. Frolov, Burial mounds near the village of Vorobyovo. *Issues Archeol. Ural* **4**, 53–87 (1962).

Acknowledgments: We would like to thank R. I. Tukhbatova, A. Immel, C.-C. Wang, and G. Brandt for laboratory assistance. We thank S. Penske and G. Neumann for sampling assistance. We thank P. Moorjani for granting access to the DATES program before publication. We are also grateful to A. V. Miller and B. K. Miller for helpful discussions on the history and archaeology of the Kazakh Steppe and to S. Reinhold for the careful revision and helpful comments on the manuscript. We thank M. O'Reilly of the graphic team of the Max Planck Institute for the Science of Human History for the graphical support. We lastly thank our colleagues of the “popgen” group of the Department of Archaeogenetics (Max Planck Institute

for the Science of Human History) W. Haak, S. Schiffels, C. Posth, C. Ning, H. Yu, K. Wang, and A. B. Rohrlach for helpful comments and discussions. **Funding:** This research was supported by the Max Planck Society and the Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean (MHAAM), the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (programs “People in the flow of history” and “History and Culture of Great Steppe” - BR05233709), the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (grant no. AP08857177), and the National Research Foundation (NRF) of Korea grant funded by the Korea government (MSIT) (2020R1C1C1003879). M.A.S. was supported by the European Research Council ERC-CoG 771234 PALEoRIDER. **Author contributions:** G.A.G.-R., C.J., L.D., and J.K. designed the study. C.J., L.D., and J.K. supervised the research. E.Kh., N.K., L.M., L.D., O.I., A.G., Z.Z., B.B., E.Ki., Z.S., A.Be., N.B., Y.B., A.Z.B., S.E., A.Bi., G.A., A.M., A.O., D.V., A.C., A.Bu., and Y.K. provided materials and resources. N.K., L.M., R.A.B., R.R., N.F.G.M., and C.F. performed genetic laboratory work. M.A.S. supervised L.M.; G.A.G.-R., C.J., L.M., M.A.S., and R.B. organized sample processing and performed preliminary quality control data analysis. G.A.G.-R. and C.J. performed population genetic data analyses. E.Kh., N.K., and L.M. collected archaeological context information, and E.Kh. organized and wrote the archaeological description with contributions from the other coauthors. G.A.G.-R. wrote the paper with contributions from C.J., J.K., E.Kh., N.K., L.D., M.A.S., and the other coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The newly produced aligned sequence data are deposited in the European Nucleotide Archive (ENA) with the following accession number: PRJEB42930. Haploid genotype data for the 1240K panel are available in eigenstrat format via the Edmond Data Repository of the Max Planck Society (<https://edmond.mpg.de>). Additional data related to this paper may be requested from the authors.

Submitted 25 August 2020

Accepted 9 February 2021

Published 26 March 2021

10.1126/sciadv.abe4414

Citation: G. A. Gneccchi-Ruscione, E. Khussainova, N. Kahbatkyzy, L. Musralina, M. A. Spyrou, R. A. Bianco, R. Radzeviciute, N. F. G. Martins, C. Freund, O. Iksan, A. Garshin, Z. Zhaniyazov, B. Bekmanov, E. Kitov, Z. Samashev, A. Beisenov, N. Berezina, Y. Berezin, A. Z. Biró, S. Évinger, A. Bissembaev, G. Akhatov, A. Mamedov, A. Onggaruly, D. Voyakin, A. Chotbayev, Y. Kariyev, A. Buzhilova, L. Djansugurova, C. Jeong, J. Krause, Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Sci. Adv.* **7**, eabe4414 (2021).