137 ancient human genomes from across the Eurasian steppe


1. Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark
2. Eco-anthropologie et Ethnobiologie, National Museum of Natural History, CNRS, Université Paris Diderot, Paris, France
3. Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, Denmark
4. Leiden University Centre for Linguistics, Leiden University, Leiden, The Netherlands
5. Department of Biology, Stanford University, Stanford, USA
6. Buketov Karaganda State University, Saryarka Archaeological Institute, Karaganda, Kazakhstan
7. Shejire DNA, Almaty, Kazakhstan
8. Department of Archaeology, Conservation and History, University of Oslo, Oslo, Norway
9. Department of Theory and Methods, Institute of Archaeology Russian Academy of Sciences, Moscow, Russia
10. Department of History, Kyrgyzstan-Turkey Manas University, Bishkek, Kyrgyzstan
11. National Academy of Sciences of Kyrgyzstan, Bishkek, Kyrgyzstan
12. Department of History, Irkutsk State University, Irkutsk, Russia
13. A.H.Margunan Institute of arhaeology, Almaty, Kazakhstan
14. Laboratory of Virology, Institute of Veterinary Medicine, Mongolian University of Life Sciences, Ulaanbaatar, Mongolia
15. Department of Biology, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar, Mongolia
16. Department of biology and ecology, Tuvan State University, Kyzyl, Russia
17. The Explico Foundation, Floro, Norway
18. Department of Archaeology, Ulaanbaatar State University, Ulaanbaatar, Mongolia
19. Department of Biology and Biotechnology, Hashemite University, Zarqa, Jordan
20. National Museum of Denmark, Unit for Environmental Archaeology and Materials Science, Kongens Lyngby, Denmark
Abstract

For thousands of years, the Eurasian steppe has been a centre for human migrations and cultural change. To understand its population history following the Bronze Age migrations, we genome-sequenced 137 ancient humans (~1X average coverage) covering the past 4000 years. We find that Scythian groups that dominated the Eurasian steppe throughout the Iron Age, were highly structured, with diverse origins comprising Late Bronze Age herders, European farmers, and South Siberian hunter-gatherers. Later, Scythians admixed with eastern steppe nomads that formed the Xiongnu confederations, and moved westward in the ~3rd/2nd century BCE, forming the Hun traditions in the 4th-5th century CE, carrying with them plague basal to the Justinian. These nomads were further admixed with East Asian groups during several short-term Medieval khanates. These historical events transformed the Eurasian steppe, from being inhabited by Indo-
European speakers of largely western Eurasian ancestry, to the present-day's mostly
Turkic-speaking groups that are primarily of East Asian ancestry.

Introduction

The Eurasian steppe stretches about 8000 kilometres from Hungary and Romania in the west, to Mongolia and northeastern China in the east. It has in the past four millennia, been dominated first by Iranian- and later Turkic- and Mongolic-speaking nomadic groups with herding and warrior economies. To understand the population genetic processes associated with the linguistic and cultural changes of the steppe following Bronze Age migrations\(^1-3\), we sequenced 137 ancient genomes to \(~\text{1X average depth}\) (see Supplementary Data Table 1 and 2) from Europe to Mongolia, and the Altai to Tian Shan mountains, covering the past \(~\text{4,000 years}\) (\(~\text{2500 BCE - 1500 CE}\)) (Figure 1). A list of the population labels used throughout the manuscript can be found in Supplementary Data Table 3. We furthermore genotyped 502 individuals of 16 self-reported ethnicities from across Central Asia, Altai, Siberia and the Caucasus (Supplementary Data Table 4 & Supplementary Section 5). In the process, we tested differential ancient DNA preservation in organic contrasted mineral substrate (Supplementary Section 6), and generated 83 new Accelerator Mass Spectrometry (AMS) dates (Supplementary Section 11).

The genomic origins of the Scythian confederations

Between \(~\text{800 and 200 BCE}\), the Eurasian steppe became dominated by the Iranian-speaking Scythians. This confederation was divided into geographically distinct groups, but was united by similarities in cultural expression\(^4\). However, the origins and population structure of the Scythians remain contested, as can be summarized in three competing models: A) the Scythians deriving from a single source originating in the northern Caucasus/steppe region\(^5-7\); B) an origin in Southern Siberia/East-Central Asia, moving westwards\(^8-9\); and finally C) the Scythians being a product of multiple transitions taking place locally, involving social and cultural borrowing in combination with gradual, small scale human movements\(^10-13\). Using PCA and admixture analyses (Figure 2 and Extended Figure 1), we observe a clear separation between two groups of Iron Age Scythians: the Hungarian Scythians and the Inner
Asian Sakas. Furthermore, we find fine-scaled structure within the Inner Asian Sakas separating: 1) the ‘Tagar’ of southern Siberia, 2) the ‘Central Sakas’ of the Central steppe, of which most have been described as belonging to the Tasmola culture (Supplementary Section 3), and 3) the 'Tian Shan Sakas' of the Tian Shan mountain range (see map in Figure 1). These differences reflect the confederal nature of the Scythian organization.

Recent genetic models suggested the presence of Yamnaya/Afanasievo ancestry in Scythians which we assessed here using a new set of outgroups that enabled us to distinguish between Early and Late Bronze Age steppe ancestry (Supplementary Section 3.6). We find that the Late Bronze Age herders are a better genetic source for the west Eurasian ancestry in Scythians than are Early Bronze Age Yamnaya/Afanasievo, the key difference being their European farmer ancestry (Supplementary Data Table 5). Using ADMIXTURE models we also illustrate the shared ancestry between Neolithic farmers (from Anatolia or Europe), Late Bronze Age herders, and Iron Age steppe nomads that is not shared with Yamnaya herders (Extended Figure 2 and Supplementary Section 5). These findings are in agreement with archaeological models.

Using D-statistics (Supplementary Section 3.7), we then characterized the sources of admixture into the various Scythian groups relatively to the Late Bronze Age steppe herders. We find that Hungarian Scythians had relatively increased European farmer ancestry (Extended Figure 3) and show no signs of Inner Asian gene flow. On the opposite, Inner Asian Sakas show relatively increased South Siberian hunter-gatherer ancestry with the strongest gene flow observed into the Central Sakas. This East Asian admixture is also reflected in the negative ‘admixture f3s’ (Extended Figure 4). We confirm the differences between these Iron Age steppe groups through D-statistics (Supplementary 3.7): the increase in Neolithic Iranian ancestry in the Tian Shan Sakas is significant when compared to Central Sakas; Tagar display increased Eastern Hunter-Gatherer (EHG) ancestry as compared to all other Scythians. Lastly, the high genetic differentiation between western and eastern Scythians is emphasized by observing higher $F_{st}$ values between Hungarian Scythians and all Inner Asian Sakas ($F_{st}$ ranges from 0.24 to 0.3) than observed among the different Inner Asian Sakas groups ($F_{st}$ ranges from 0.15 to 0.2) (Supplementary Data Table 6).

The qpAdm modelling of this ancient genomic dataset is consistent with these findings. The Central Sakas can be modelled as a simple two-way mixture of Late Bronze Age pastoralists.
and South Siberian hunter-gatherers (BHG_BA), with almost equal proportions of Bronze Age herder (56%) and South Siberian hunter-gatherer ancestry (44%). The southern Siberian Tagar show unequal ancestry contributions from Bronze Age herders (83.5%) and South Siberian hunter-gatherers (7.5%), but also an additional contribution of Mal'ta (MA1)-like ancestry (9%) indicating differences in the sources of hunter-gatherer admixture across the Sakas. The Saka population of the Tian Shan mountains displays a high proportion of Late Bronze Age steppe herder ancestry (70%) followed by South Siberian hunter-gatherer ancestry (25%), but also reveals an additional 5% ancestry coming from a source related to a Neolithic population from Iran. Taken together, our data do not support the recent mtDNA-based claim of extensive gene flow between the different Scythian groups, but indicate instead admixture between populations of Late Bronze Age herder descent and various local groups, in agreement with the multiple origins model (C).

Altogether, our data show that the culturally similar Scythians represented genetically structured groups within the Eurasian steppe. In particular, the Siberian Tagar, Central Sakas, and the Tian Shan Sakas were Scythian groups that arose through admixture between Late Bronze Age pastoral groups and Inner Asian hunter-gatherers, in contrast to the Hungarian Scythians that received gene flow from farming groups within Europe. The additional gene flow from a source related to the Neolithic Iranians detected in the Tian Shan Sakas suggests that southern steppe nomads also interacted with the civilization of the Bactria-Margiana Archaeological Complex (BMAC) of present-day eastern Turkmenistan.

The Xiongnu and the Hunnic expansions

Turkic language elements arguably first emerged among the Xiongnu nomads, a confederation of several nomadic tribes who occupied the eastern steppe from the 3rd century BCE. They are believed to be of East Asian ancestry, although ancient Y-chromosomal data have indicated a possibly heterogeneous population admixed with central steppe nomads. Huns (3rd-5th century CE) have been argued to derive directly from the Xiongnu while others claim that there is no evidence connecting the two groups. It is commonly believed that the Huns spread westward, disseminating Turkic languages throughout Central Asia at the cost of Iranian languages. It is known that the expansion of the Xiongnu nomads impacted the movements of other cultural groups from the south-eastern side of the Tian Shan Mountains, such as the Wusun and Kangju, whose genetic ancestries have so far remained
unknown. Based on the archaeological record it has tentatively been suggested that they belonged to the Iranian-speaking branch of the Indo-European language family. Principal Component Analyses and D-statistics suggest that the Xiongnu individuals belong to two distinct groups, one being of East Asian origin and the other presenting considerable admixture levels with West Eurasian sources (Figure 2, Extended Figures 1 and Extended Figure 5, respectively labelled “Xiongnu” and “Xiongnu_WE”). We find that Central Sakas are accepted as a source for these ‘western-admixed’ Xiongnu in a single-wave model. In line with this finding, no East Asian gene flow is detected compared to Central Sakas as these form a clade with respect to the East Asian Xiongnu in a D-statistic, and furthermore, cluster closely together in the PCA (Figure 2).

We used D-statistics (Supplementary Section 3.7) to investigate the genetic relationship between Iron Age nomads, the East Asian Xiongnu, and the early Huns of the Tian Shan. We find that the Huns have increased shared drift with West Eurasians as compared to the Xiongnu (Extended Figure 6). We tested for patterns of shared drift between the Xiongnu and the Wusun, the preceding Sakas, and the slightly later Huns (2nd century CE). We find that both the earlier Sakas and the later Huns have more East Asian ancestry than the Wusun. This is also apparent from model-based clustering and PCA (Extended Figure 7). Similar results are seen with the contemporaneous and later Kangju groups that, like the Wusun, re-emerged into the central steppe from south-east of the Tian Shan mountains. In addition, both groups require a Neolithic Iranian-related source for modelling ancestral proportions in the qpAdm framework (Supplementary Data Table 7), together with Late Bronze Age pastoralists and the South Siberian hunter-gatherers. We therefore suspect that the Wusun and Kangju groups are descendants of Bronze Age pastoralists that interacted with the BMAC civilization in southern Uzbekistan/eastern Turkmenistan, yet remained much less admixed with East Asians than did the Iron Age steppe Sakas.

Overall, our data show that the Xiongnu confederation was genetically heterogeneous, and that the Huns emerged following minor male-driven East Asian gene flow into the preceding Sakas that they invaded (see Supplementary Section 3.6 for sex-biased admixture rates). As such our results support the contention that the disappearance of the Inner Asian Scythians/Sakas around the beginning of the Common Era was a cultural transition that coincided with the westward migration of the Xiongnu. This Xiongnu invasion also led to the
displacement of isolated remnant groups related to Late Bronze Age pastoralists living on the south-eastern side of the Tian Shan mountains.

Repeated conquests and waves of East Asian impact

In the 6th Century CE the Hunnic Empire had been broken up and dispersed, as the Turk Khaganate overtook the military and political domination of the steppe. Khaganates were steppe nomad political organizations that varied in size, and became dominant during this period, and can be contrasted to the previous stateless organization of the Iron Age. The Turk Khaganate was eventually replaced by a number of short-lived steppe cultures. These included the Kipchak and the Tungusic Kimak populations spreading southwards towards the Tian Shan and westward towards the Ural Mountains forming the Kimak Khaganate of the central steppe during the 8th to 11th Century CE. Towards the 11th Century, the Kimak Khaganate was overthrown by local Kipchak groups, who in turn allied themselves with the Cuman of western Eurasia. Eventually, the short-lived Khaganates were overtaken by the Mongol Empire that emerged through unification of East Mongolian and Transbaikalian tribes that expanded heavily during Genghis Khans' rule in the 13th Century CE.

We find evidence that elite soldiers associated with the Turk Khaganate are genetically closer to East Asians than are the preceding Huns of the Tian Shan mountains. We also find that one Turk period nomad was a genetic outlier with pronounced European ancestries, indicating the presence of ongoing contact with Europe. Only one sample here represents Kimak nomads, and it does not show elevated East Asian ancestry. During the Kipchak period in the 11th Century CE, the rule was allegedly taken over by another group originating from the geographical area of Tuva. We present genomic data from two individuals from this period, one of which shows increased East Asian ancestry, while the other has pronounced European ancestry. These individuals date back to the Cuman-Kipchak alliance, which englobed both the western and eastern steppe. For the period when the region became incorporated into the Karakhanid Khanate that encompassed present-day regions of Uzbekistan, Tajikistan, Kazakhstan and Kyrgyzstan, D-statistics identify a small influx of East Asian ancestry as compared to the earlier Turk period. In agreement with this, nomads in the Karakhanid period are shifted towards East Asians when compared to earlier Turks in the...
PCA plot (Figure 2 and Extended Figure 8). Additionally, we analyzed 10 culturally
unaffiliated Medieval nomads, most showing pronounced East Asian ancestry, albeit at very
different proportions (Extended Figure 8). We also find the presence of an individual of
western European descent buried together with members of Jochi Khan’s Golden Horde army
from the Ulytau mountains (Supplementary Section 4 – ‘DA28’ is ‘East Asian’ and ‘DA29’ is
‘European’). This could suggest assimilation of distinct groups into the Medieval Golden
Horde, but this individual may also represent a slave or a servant of west Eurasian descent
attached to the service of the Golden Horde members.

These results suggest that Turk cultural customs were imposed by an East Asian minority elite
onto central steppe nomad populations, resulting in a small detectable increase in East Asian
ancestry. However, we also find that steppe nomad ancestry in this period was extremely
heterogeneous with several individuals being genetically distributed at the extremes of the
first principal component (Figure 2) separating Eastern and Western descent. Based on this
notable heterogeneity, we interpret that during Medieval times, the steppe populations were
exposed to gradual admixture from the East, while interacting with incoming west Eurasians.
The strong variation is a direct window into ongoing admixture processes and to the multi-
ethnic cultural organization of this period.

Origins and spread of the Justinian plague

A few decades after the period of Hunnic driven mobility across the Eurasian steppe, large
areas of Europe were depopulated due to the Justinian plague pandemic\(^{28}\). While the first
reports of the pandemic point to an outbreak in Egypt from where it is thought to have spread
into Europe\(^{29}\), the primordial origins of the Justinian plague remain unknown. The most basal
strains of present-day plague (0.PE7 clade) have been found in Qinghai, south-east of the Tian
Shan mountains\(^{30}\), and the clade basal to the Justinian plague (0.ANT1) was found in Xinjiang
in China, thus pointing to a possible Inner Asian origin of the Justinian plague.

We find that two individuals show detectable levels of \(Y.\) pestis DNA, compatible with the
characterization of the full genome sequence at 8.7X and 0.24X coverage, respectively. The
first individual is a Hun from the Tian Shan Mountains (DA101), dating to \(~180\) CE, while
the second individual is from the Alan culture from North Ossetia (DA147), dating \(6^{th}-9^{th}\)
century CE. The genome from the \(Y.\) pestis strain DA101 which we name 0.ANT5, branches
off from the main plague lineage just basal to the Justinian plague strain 0.ANT4, identified from an individual in Aschheim, Germany and dated to ~530 CE\(^2\) (Extended Figure 9). As expected the Tian Shan strain contained the \textit{ymt} gene reported to be missing in the more ancestral Bronze Age plague strains\(^3\). The strain also displayed the loss of function mutations in \textit{pde2}, \textit{pde3}, \textit{rcsA} and \textit{ureD} that are required for flea transmission in the traditional blocked flea model\(^4\) (Extended Figure 9). This, coupled with a fully functional plasminogen activator gene, indicates that the ‘Hunnic’ plague strain had full bubonic capability and flea transmissibility.

We find a higher number of strain-specific variants in the Aschheim strain is in line with the difference in sampling time (~180 CE vs. ~530 CE) and the potentially multiple replication cycles associated with pandemics\(^3\). This is supported by the substitution rate on the branch leading to the Aschheim strain being higher. Mutation rates in pathogens have been hypothesized to be affected by epidemics; not only because of natural selection, but also due to an increase in replication rate\(^3\). Therefore, our observation of an accelerated mutation rate is in agreement with this hypothesis and supports that the Ascheim strain was responsible for a major outbreak – the Justinian plague.

Given that the most basal strains of present-day plague (0.PE7 clade) originate from Qinghai\(^3\), and the clade basal to the Justinian plague (0.ANT1) is from Xinjiang, China, two areas close to the Tian Shan, we find provisional support for the hypothesis that the pandemic was brought to Europe towards the end of the Hunnic period through the Silk Road along the southern fringes of the steppe.

**Discussion**

The overall population history that formed the genetic composition of present-day steppe populations is illustrated in Figure 3, where we model the entire known ancient and present-day diversity of Inner Asia using the key ancestral groups. We also identify sex-specific admixture proportions in the Iron Age (Extended Figure 10 and Supplementary Section 3.6). In Figure 4 we present main migratory patterns. Our findings fit well with the current insights in the historical linguistics of this region (Supplementary Section 2). The steppe was likely largely Iranian-speaking in the 2\(^{nd}\) and 1\(^{st}\) millennia BCE. This is supported by the split of the Indo-Iranian linguistic branch into Iranian and Indian\(^3\), the distribution of the Iranian
languages, and the preservation of Old Iranian loanwords in Tocharian. The wide
distribution of the Turkic languages from Northwest China, Mongolia and Siberia in the east
to Turkey, Bulgaria, Romania and Lithuania in the west implies large-scale migrations out of
the homeland in Mongolia since the beginning of the Common Era. The diversification
within the Turkic languages suggests that several waves of migrations occurred, and on the
basis of the impact of local languages gradual assimilation to local populations were already
assumed. The East Asian migration starting with the Xiongnu complies well with the
hypothesis that early Turkic was their major language. Further migrations of East Asians
westwards find a good linguistic correlate in the influence of Mongolian on Turkic and
Iranian in the last millennium. As such, the genomic history of the Eurasian steppe is the
story of a gradual transition from Bronze Age pastoralists of western Eurasian ancestry,
towards mounted warriors of increased East Asian ancestry – a process that continued well
into historical times.

Data availability
Sequence data were deposited in the European Nucleotide Archive (ENA) under accession:
PRJEB20658 (ERP022829). SNP data for present-day populations are available after ethical
validation in the European Genome-Phenome Archive (EGA) under accession:
EGAS00001002926.

References
1. Haak, W. et al. Massive migration from the steppe was a source for Indo-European
4. Chlenova, N. L. On the Degree of Similarity between Material Culture Components


Acknowledgments
We thank Kim Magnussen, Lillian Petersen, Cecilie Mortensen and Andaine Seguin-Orlando at the Danish National Sequencing Centre for producing the analysed sequences. We thank Paula Reimer and Stephen Hoper at the 14Chrono Center Belfast for providing AMS datings. We thank Susanne Hackenbeck for discussing paleodietary reconstructions. We thank Ditte Christiansen Appelt, Helene Elisabeth Heyerdahl, the Explico Foundation team, Jainagul Isakova, Batyrzhan Daulet, Aidyn Taиров, Nurlan Abduov, Bakhtishat Tudiyarov, Vladimir Volkov, Maksum Akchurin, Ilyas Baimukhan, Nikolay Namdakov, Yuldash Yusupov, Erlan Ramankulov, Arman Nurgaziyev, Abdul Kusaev for important assistance in fieldwork. We thank Jesper Stenderup, Pernille V. Olsen and Tina Brand for technical assistance in the laboratory. We thank all involved archaeologists, historians and geographers from Kazakhstan: Alexander Suslov, Irina Erofeeva, Erzhan Nurmaganbetov, Baktyar Kozhakhmetov, Nadezhda Loman, Yuri Parshin, Sergey Ladunskiy, Marina Bedelbaeva, Antónia Marcšik, Oliver Gábor, Marek Půlpán, Yerkin Kubeev, Rymbek Zhumashev, Khylıshe Omarov, Serik Kasymov, Umut Akimbayeva. We thank Paul Rodzianko for creating the initial contact between PBD, SE and EU. We thank Sten Jacobsen and Jonh O'Brien for translating and proof-reading Russian contributions. The project was funded by the Danish National Research Foundation (EW), the Lundbeck Foundation (EW), and KU2016 (EW).

Author contributions:

EW: initiated and led study
PBD, EW, EU, EH: designed study
PBD and NM: produced the data
PBD, NM, SR, MS, GR, ThK, AG, MWP, AGP, KN: analysed or assisted in analysis of data
PBD, EW, KK: interpreted results with considerable input from MS, RN, MP, NK, SR, LO,
MEA, JVMM
PBD, EW, KK, MP, SR: wrote the manuscript with considerable input from NK, LH, MS,
RN, MEA, LO, JVMM and contributions from all authors
PBD, MEA, LO, EU, NB, VL, GA, KA, AlA, AsA, GB, VIB, AB, BazartserenBoldbaatar,
BazartserenBoldgiv, CD, SE, DE, RD, ED, VE, KF, AG, AG, HH, TH, ZK, RK, EK, AK,
TaK, AK, IK, NL, AM, VKM, IVM, IM, EM, VM, GM, BN, ZO, IP, KP, VS, IS, AL, KGS,
TS, KT, AT, TT, DV, LY, SU, VV, AW, EH excavated, curated, sampled and/or described
analysed skeletons
All authors contributed to final interpretation of data.

**Competing financial interests.**
The authors declare no competing financial interests.

**Figure Captions**

**Figure 1. Cultural and geographical presentation of the ancient samples.** a) Geographical distribution of samples. Symbols corresponds to samples of a specific age; circle: Bronze age, square: Iron Age, diamond: Hun period, triangle upwards: Turk period, triangle downwards: Medieval times. b) Each symbol has been sorted according to geographical region highlighted on map (panel A), and are given in the grey boxes in panel b. Abbreviations corresponds C = Caucasus, CAS = Caspian steppe, CS = Central steppe, ES = Eastern steppe, HP = Hungarian plains, PS = Pontic steppe, STE = Siberia, Tungus & Eastern Steppe.

**Figure 2. Principal Component Analyses.** The Principal Components 1 and 2 were plotted for the ancient data analysed with the present-day data (no projection bias) using 502 individuals at 242,406 autosomal SNP positions. Dimension 1 explaining 3% of the variance represents a gradient stretching from Europe to East Asia. Dimension 2 explaining 0.6% of the variance, a gradient mainly represented by ancient DNA starting from 'basal-rich' cluster of Natufian hunter-gatherers and ending with Eastern Hunter-Gatherer (EHG).

**Figure 3. QpAdm results depicts the changes in ancestry across time in Central Asia.** The changes reflect a gradual increase in East Asian ancestry in the central steppe nomads coupled to a decrease in ancestry associated to Eastern-Hunter Gatherers, starting high in Yamnaya and finishing low in present-day Kazakhs/Kyrgyz. The set of outgroups used is: Mbuti, Ust'Ishim, Clovis, Kostenki14 and Switzerland HG.

**Figure 4. Summary review map.** Depiction of the five main migratory events associated to the genomic history of the steppe pastoralists, 3000 BCE – present. a) A depiction of Early Bronze Age migrations related to the expansion of Yammaya and Afasievo culture, b) A depiction of Late Bronze Age migrations related to the Sintashta and Andronovo horizons, c)
A depiction of Iron Age migrations and sources of admixture, d) A depiction of Hun period migrations and sources of admixture, e) A depiction of Medieval migrations across the steppe.

Extended Figure 1. Analyses of Iron Age clusters. A) PCA of Iron Age nomads and ancestral sources explaining the diversity between them using 74 individuals at 242,406 autosomal SNP positions. B) PCA of Iron Age nomads alone using 29 individuals at 242,406 autosomal SNP positions. C) PCA of Xiongnu, 'Western' Xiongnu, Tian Shan Huns, Nomads Hun Period, and Tian Shan Sakas thus 39 individuals at 242,406 autosomal SNP positions. E) Results for model-based clustering analysis at K=7. Here we illustrate the admixture analyses with K=7 as it approximately identifies the major component of relevance (Anatolian/European farmer component, Caucasian ancestry, EHG related ancestry and East Asian ancestry). D) Model-based clustering at K=7 illustrating differences in ancestral proportions – individuals labelled (A) Andronovo, (B) Neolithic European (Europe_EN), (C) East Asian Hunter-Gatherer (BHG_BA), (D) Neolithic Iranian (Iran_N). Here we illustrate the admixture analyses with K=7 as it approximately identifies the major component of relevance (Anatolian/European farmer component, Caucasian ancestry, EHG related ancestry and East Asian ancestry). The * symbols indicate individuals flagged as genetic outliers. E) Results for model-based clustering analysis at K=7. Here we illustrate the admixture analyses with K=7 as it approximately identifies the major component of relevance (Anatolian/European farmer component, Caucasian ancestry, EHG related ancestry and East Asian ancestry).

Extended Figure 2. Illustration of shared ancestry between Neolithic farmers and Iron Age nomads. Results for model-based clustering analysis at K=7, plotting only one individual from relevant groups, in order to illustrate shared ancestry between Neolithic farmers from Anatolia/Europe, Late Bronze Age nomads and Iron Age nomads not shared with Early Bronze Age nomads.

Extended Figure 3. Illustration of gene flow into Hungarian Scythians. We represent all D (Test, Mbuti; Andronovo, Hungarian Scythians) that deviate significantly from 0 (i.e. higher than 3 times the standard errors). The reported numbers are the D-statistics and the 3 standard errors were plotted as error bars. The number of individuals per population can be found in Supplementary Data Table 3 and 4.
Extended Figure 4. Illustration of negative admixture f3 statistics for Iron Age populations. \( f_3(\text{Bronze Age Test1, Bronze Age Test2; Iron Age Test}) \). The reported numbers of the f3 statistics and 3 standard errors were plotted as errors bars. The number of individuals per population can be found in Supplementary Data Table 3.

Extended Figure 5. Illustration of West Eurasian geneflow into groups forming the Xiongnu culture. We represent all \( D(\text{Test, Mbuti; 'Western' Xiongnu, Xiongnu}) \) that deviate significantly from 0 (i.e. higher than 3 times the standard errors). The reported numbers are the D-statistics and the 3 standard errors were plotted as error bars. The number of individuals per population can be found in Supplementary Data Table 3 and 4.

Extended Figure 6. Illustration of West Eurasian ancestry in early Tian Shan Huns. We represent all \( D(\text{Test, Mbuti; Tian Shan Huns, Xiongnu}) \) that deviate significantly from 0 (i.e. higher than 3 times the standard errors). The reported numbers are the D-statistics and the 3 standard errors were plotted as error bars. The number of individuals per population can be found in Supplementary Data Table 3 and 4.

Extended Figure 7. Analyses of Xiongnu and Hun period population clusters. A) PCA of Xiongnu, 'Western' Xiongnu, Tian Shan Huns, Nomads Hun Period, and Tian Shan Sakas, Kangju and Wusun thus 49 individuals at 242,406 autosomal SNP positions. B) Results for model-based clustering analysis at K=7. Here we illustrate the admixture analyses with K=7 as it approximately identifies the major component of relevance (Anatolian/European farmer component, Caucasian ancestry, EHG related ancestry and East Asian ancestry). (A) is a South Siberian individual associated to Andronovo culture.

Extended Figure 8. Analyses of Turk and Medieval period population clusters. A) PCA of Tian Shan Hun, Turk, Kimak, Kipchack, Karakhanid and Golden Horde thus 28 individuals at 242,406 autosomal SNP positions. B) Results for model-based clustering analysis at K=7. Here we illustrate the admixture analyses with K=7 as it approximately identifies the major component of relevance (Anatolian/European farmer component, Caucasian ancestry, EHG related ancestry and East Asian ancestry).

Extended Figure 9. Maximum Likelihood phylogenetic reconstruction of \( Y. \ pestis \). This tree reveals the basal position of the Tian Shan sample (0.ANT5, DA101, 186 CE) compared
to the Justinian plague sample (0.ANT4: A120, 536 CE). These two samples are shown in
orange italics. Other ancient plague samples included in the tree are the Bronze Age samples
(0.PRE1 and 0.PRE2) and a Black Death sample (1.PRE1). Numbers on nodes indicate
bootstrap support (not all shown for clarity) and certain branches have been collapsed for
clarity. Branch lengths are substitutions per site.

**Extended Figure 10. Analyses of sex-specific contributions to Iron Age populations.**

Estimates of the male and female contributions from each source populations (left column) to
each of the four admixed populations (right column) using the method of Goldberg et al.
(2017). For each admixed population, we compared the observed mean autosomal and X-
chromosomal ancestry, estimated in qpAdm, to that calculated under a constant admixture
model on a grid of sex-specific contribution parameters ranging from 0 to 1 in 0.025
increments using a Euclidean distance. The log of the ratio of male to female contribution
parameters that produce the smallest 0.1% of distances from the data are plotted, with the full
range of parameter values in grey, the middle 50% in black, and the median value in red. The
dashed line indicates equal male and female contributions.

**Supplementary Data Table 1 – Basic mapping statistics.**

**Supplementary Data Table 2 – Overview of ancient samples.** This table includes
radiocarbon dating and calibration, geographical coordinates and genetic gender.

**Supplementary Data Table 3 – Population label and sample size overview.** This table
provides a fast contextualization of population labels used here.

**Supplementary Data Table 4 – Information on present-day dataset.** This includes
geographical coordinates and sample size coupled to the full presentation of ancestral
proportions estimated using qpAdm with a set of 5 outgroups: Mbuti, Ust'Ishim, Clovis,
Kostenki14 and Switzerland HG. See Supplementary Section 3 for description of qpAdm
analyses.

**Supplementary Data Table 5 – QpAdm modelling of Iron Age Scythians.** We here
compare different sets of sources, ie. Andronovo, Sintashta and Yamnaya and a set of 7
outgroups (Mbuti, Ust'Ishim, Clovis, Kostenki14, Switzerland_HG, Natufian and MA1). Red colors reflect a failed model. Note that for Tagar where MA1 was used a source, the outgroup was replaced with EHG. Number of individuals per modelled population can be found in Supplementary Data Table 3. See Supplementary Section 3 for description of qpAdm analyses.

Supplementary Data Table 6 – F_{st} values between the Iron Age Scythian groups. Number of individuals per modelled population can be found in Supplementary Data Table 3.

Supplementary Data Table 7 - QpAdm modelling of Kangju and Wusun. We here use a set of 7 outgroups (Mbuti, Ust'Ishim, Clovis, Kostenki14, Switzerland_HG, Natufian and MA1). Number of individuals per modelled population can be found in Supplementary Data Table 3. See Supplementary Section 3 for description of qpAdm analyses.

Supplementary Data Table 8 – Authentication assessment. Damage parameters, contamination estimates and mitogenome haplogroup assignment. See Supplementary Section 3 and 10 for exhaustive description of sample analyses.

Supplementary Data Table 9 – Confident Y-chromosomal haplogroup assignment.
Early Bronze Age steppe
Middle Bronze Age steppe
Iron Age steppe
'Medieval' steppe
Kazakhs
Central Steppe

Present-day

Medieval Period

Hun Period

Iron Age

Middle to late Bronze Age

Early Bronze Age

Pontian Steppe

Caspian Sea

Altai Mt.

Tian Shan

East Asian Hunter Gatherer

Western Hunter Gatherer

Eastern Hunter Gatherer

Natufian

50N

40N

Europe

A

A

50E

80E

90E

70E

Iron Age

Early Bronze Age

Middle to late Bronze Age

Hun Period

Medieval Period

Present-day
3000 - 2100 BCE

2100 - 1200 BCE

1200 - 200 BCE

200 BCE - 600 CE

600 - 1500 CE

Yamnaya/Afanasievo expansion
Sintashta, Srubnaya, Andronovo expansion
Scythian expansion
Late Bronze Age admixture
Xiongnu - Hunnic expansion
Asian medieval impact