137 ancient human genomes from across the Eurasian steppes

Barros Damgaard, P. de; Marchi, N.; Rasmussen, S.; Peyrot, M.; Renaud, G.; Korneliussen, T.; ... ; Willerslev, E.

Citation

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

ARTICLE

https://doi.org/10.1038/s41586-018-0094-2

137 ancient human genomes from across the Eurasian steppes

Peter de Barros Damgaard[1], Nina Marchi[2], Simon Rasmussen[3], Michaël Peyrot[4], Gabriel Renaud[1], Thorfinn Korneliussen[1,5], J. Víctor Moreno-Mayar[1], Mikkel Winther Pedersen[5], Amy Goldberg[6], Emma Usmanova[7], Nurbol Baimukhanov[8], Valeriy Loman[7], Lotte Hedeager[9], Anders Gorm Pedersen[3], Kasper Nielsen[3,51], Gennady Afanasiev[10], Kunbolot Akmatov[11], Almaz Aldashev[12], Ashyk Alpaslan[11], Gabit Baimbetov[8], Vladimir I. Bazaliiskii[13], Arman Beisenov[14], Bazartseren Boldbaatar[15], Bazartseren Boldgiv[16], Choduraa Dorzhu[17], Sturla Ellingvag[18], Diimaajav Erdenebaatar[19], Rana Dajani[20,21], Evgeniy Dmitriev[7], Valeriy Evdokimov[7], Karin M. Frei[22], Andrey Gromov[23], Alexander Goryachev[24], Hakon Hakonarson[25], Tatyana Hegay[26], Zaruhi Khachatryan[27], Ruslan Khaskhanov[28], Egor Kitov[14,29], Alina Kolbina[30], Tabaldiev Kubatbek[11], Alexey Kukushkin[7], Igor Kukushkin[7], Nina Lau[31], Ashot Margaryan[1,32], Inga Merkyte[33], Ilya V. Mertz[34], Viktor K. Mertz[34], Enkhbayar Mijiddorj[19], Vyacheslav Moiyesev[23], Gulmira Mukhtarova[35], Bekmukhanbet Nurmukhanbetov[35], Z. Orozbekova[36], Irina Panyushkina[37], Karol Pieta[38], Václav Smrčka[39], Irina Shevnina[40], Andrey Logvin[40], Karl-Göran Sjögren[41], Tereza Štolcová[38], Angela M. Taravella[50], Kadicha Tashbaeva[42], Alexander Tkachev[43], Turaly Tulegenov[35], Dmitriy Voyakin[24], Levon Yepiskoposyan[27], Sainbileg Undrakhbold[16], Victor Varfolomeev[7], Andrzej Weber[44], Melissa A. Wilson Sayres[50], Nikolay Kradin[45,46], Morten E. Allentoft[1], Ludovic Orlando[1,47], Rasmus Nielsen[1,48], Martin Sikora[1], Evelyne Heyer[2], Kristian Kristiansen[41] & Eske Willerslev[1,5,49]*

For thousands of years the Eurasian steppes have been a centre of human migrations and cultural change. Here we sequence the genomes of 137 ancient humans (about 1× average coverage), covering a period of 4,000 years, to understand the population history of the Eurasian steppes after the Bronze Age migrations. We find that the genetics of the Scythian groups that dominated the Eurasian steppes throughout the Iron Age were highly structured, with diverse origins comprising Late Bronze Age herders, European farmers and southern Siberian hunter-gatherers. Later, Scythians admixed with the eastern steppe nomads who formed the Xiongnu confederations, and moved westward in about the second or third century BC, forming the Hun traditions in the fourth–fifth century AD, and carrying with them plague that was basal to the Justinian plague. These nomads were further admixed with East Asian groups during several short-term khanates in the Medieval period. These historical events transformed the Eurasian steppes from being inhabited by Indo-European speakers of largely West Eurasian ancestry to the mostly Turkic-speaking groups of the present day, who are primarily of East Asian ancestry.

The Eurasian steppes stretch about 8,000 km from Hungary and Romania in the west to Mongolia and northeastern China in the east. These regions have, in the past four to five millennia, been dominated first by Iranian- and later by 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 steppes after the Bronze Age migrations[1–3], we sequenced 137 ancient genomes—to about 1× average depth (see Supplementary Tables 1, 2)—from Europe to Mongolia and the Altai to Tian Shan mountains; these genomes covered approximately 4,000

---

[1]Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [2]Eco-anthropologie et Ethnobiologie, Muséum national d'Histoire naturelle, 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 Zoology, University of Cambridge, Cambridge, UK. [6]Department of Biology, Stanford University, Stanford, CA, USA. [7]Buketov Karaganda State University, Saryarka Archaeological Institute, Karaganda, Kazakhstan. [8]Shejire DNA, Almaty, Kazakhstan. [9]Department of Archaeology, Conservation and History, University of Oslo, Oslo, Norway. [10]Department of Theory and Methods, Institute of Archaeology Russian Academy of Sciences, Moscow, Russia. [11]Department of History, Kyrgyzstan-Turkey Manas University, Bishkek, Kyrgyzstan. [12]National Academy of Sciences of Kyrgyzstan, Bishkek, Kyrgyzstan. [13]Department of History, Irkutsk State University, Irkutsk, Russia. [14]A. Kh. Margulan Institute of Archaeology, Almaty, Kazakhstan. [15]Laboratory of Virology, Institute of Veterinary Medicine, Mongolian University of Life Sciences, Ulaanbaatar, Mongolia. [16]Department of Biology, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar, Mongolia. [17]Departament of Biology and Ecology, Tuvan State University, Kyzyl, Russia. [18]The Explico Foundation, Floro, Norway. [19]Department of Archaeology, Ulaanbaatar State University, Ulaanbaatar, Mongolia. [20]Department of Biology and Biotechnology, Hashemite University, Zarqa, Jordan. [21]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA, USA. [22]Unit for Environmental Archaeology and Materials Science, National Museum of Denmark, Copenhagen, Denmark. [23]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, St. Petersburg, Russia. [24]Archaeological Expertise LLC, Almaty, Kazakhstan. [25]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. [26]Republican Scientific Center of Immunology, Ministry of Public Health, Tashkent, Uzbekistan. [27]Department of Bioengineering, Bioinformatics and Molecular Biology, Russian-Armenian University, Yerevan, Armenia. [28]Complex Research Institute of the Russian Academy of Sciences, Grozny, Russia. [29]Institute of Ethnology and Anthropology, Russian Academy of Science, Moscow, Russia. [30]Kostanay Regional Local History Museum, Kostanay, Kazakhstan. [31]Centre for Baltic and Scandinavian Archaeology, Schleswig, Germany. [32]Laboratory of Ethnogenomics, Institute of Molecular Biology, National Academy of Sciences of Armenia, Yerevan, Armenia. [33]Saxo-Institute, University of Copenhagen, Copenhagen, Denmark. [34]Center for Archaeological Research, S. Toraighyrov Pavlodar State University, Pavlodar, Kazakhstan. [35]The State Historical and Cultural Reserve-Museum (ISSYK), Almaty, Kazakhstan. [36]Institute of Archeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia. [37]University of Arizona, Laboratory of Tree-Ring Research, Tucson, AZ, USA. [38]Institute of Archaeology of the Slovak Academy of Sciences, Nitra, Slovakia. [39]Institute for History of Medicine and Foreign Languages, First Faculty of Medicine, Charles University, Prague, Czech Republic. [40]Archaeological Laboratory, Kostanay State University, Kostanay, Kazakhstan. [41]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden. [42]Institute of History and Cultural Heritage of National Academy of Sciences, Bishkek, Kyrgyzstan. [43]Institute of Problems Development of the North Siberian Branch of the Russian Academy of Sciences, Tyumen, Russia. [44]Department of Anthropology, University of Alberta, Edmonton, Alberta, Canada. [45]Institute of History, Archaeology and Ethnology, Far-Eastern Branch of the Russian Academy of Sciences, Ulan-Ude, Russia. [46]Institute of Mongolian, Buddhist, and Tibetan Studies, Siberian Branch of the Russian Academy of Sciences, Ulan-Ude, Russia. [47]Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, Université de Toulouse, Université Paul Sabatier, Toulouse, France. [48]Departments of Integrative Biology and Statistics, University of Berkeley, Berkeley, CA, USA. [49]Wellcome Trust Sanger Institute, Hinxton, UK. [50]School of Life Sciences, Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, AZ, USA. [51]Present address: Carlsberg Research Laboratory, Copenhagen, Denmark. *e-mail: ewillerslev@snm.ku.dk

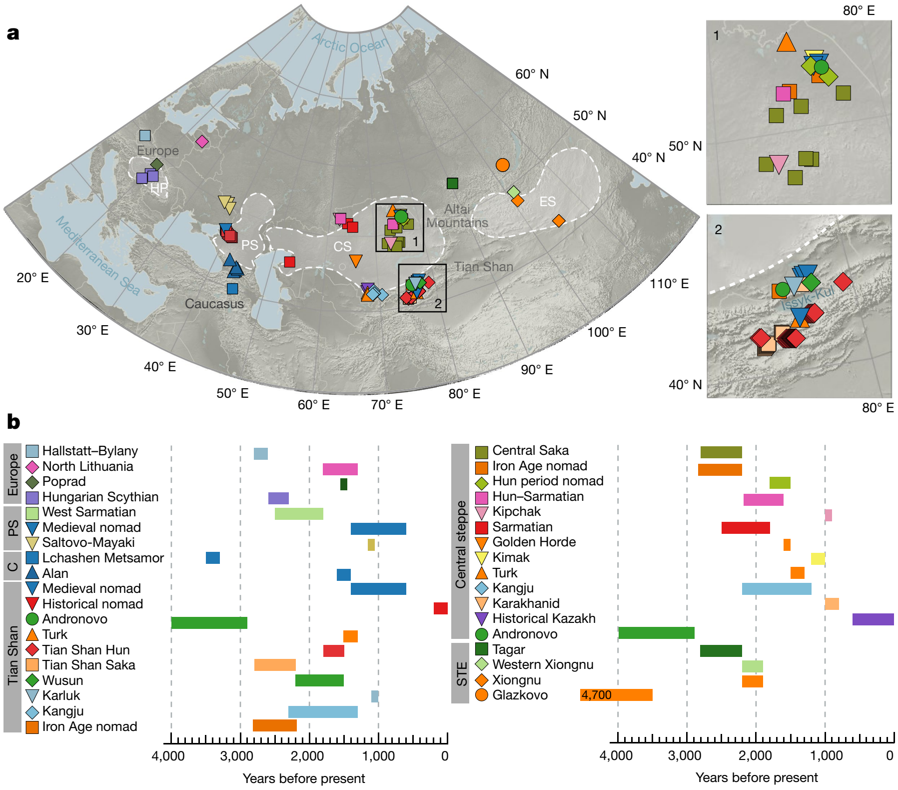


Fig. 1 | Cultural and geographical presentation of the ancient samples. **a**, Geographical distribution of samples. Symbols correspond to samples of a specific age: circle, Bronze Age; square, Iron Age; diamond, Hun period; triangle upwards, Turk period; triangle downwards, Medieval period. **b**, Each symbol has been sorted according to geographical region highlighted on the map in **a**, and given in the grey boxes in **b**. C, Caucasus; CS, central steppe; ES, eastern steppe; HP, Hungarian plains; PS, Pontic steppe; STE, Siberia, Tungus and eastern steppe (STE is not marked on the map in **a**, but includes steppe and non-steppe areas).

years (about 2500 BC–AD 1500) (Fig. 1). A list of the population labels used throughout this Article can be found in Supplementary Table 3. Furthermore, we genotyped 502 individuals of 16 self-reported ethnicities from across Central Asia, Altai, Siberia and the Caucasus (Supplementary Table 4 and Supplementary Information section 5). In the process, we tested differential ancient DNA preservation in organic contrasted mineral substrate (Supplementary Information section 6), and generated 83 new accelerator mass spectrometry dates (Supplementary Information section 11).

The genomic origins of the Scythian confederations

Between about 800 and 200 BC, the Eurasian steppes 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: (1) the Scythians deriving from a single source originating in the northern Caucasus or steppe region[5–7]; (2) an origin in southern Siberia or east-central Asia, moving westwards[8,9]; and finally (3) 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 principal component analysis (PCA) and ADMIXTURE analyses (Fig. 2 and Extended Data Fig. 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 that separates (1) the populations associated with the 'Tagar' culture of southern Siberia, (2) the 'Central Sakas' of the central steppe—most of whom have been described as belonging to the Tasmola culture (Supplementary Information section 3)—and (3) the 'Tian Shan Sakas' of the Tian Shan mountain range (see map in Fig. 1). These differences reflect the confederal nature of the Scythian organization.

Recent genetic models suggested the presence of Yamnaya and/or Afanasievo ancestry in Scythians[11], which we assessed here using a new set of outgroups that enabled us to distinguish between Early and Late Bronze Age steppe ancestry (Supplementary Information 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 or Afanasievo, the key difference being their European farmer ancestry (Supplementary Table 5). Using ADMIXTURE models[14] 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 Data Fig. 2 and Supplementary Fig. 163). These findings are consistent with archaeological models.

Using *D*-statistics (Supplementary Information section 3.7), we then characterized the sources of admixture into the various Scythian groups relative to the Late Bronze Age steppe herders. We find that Hungarian Scythians had relatively increased European farmer ancestry (Extended Data Fig. 3) and show no signs of gene flow from Inner Asian groups. Conversely, Inner Asian Sakas show relatively increased southern 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 f_3 values, indicating that Late Bronze Age pastoralists and southern Siberian hunter-gatherers are excellent proxies for the admixing populations (Extended Data Fig. 4). We confirm the differences between these Iron Age steppe groups through *D*-statistics (Supplementary Information section 3.7). The increase in Neolithic Iranian ancestry in the Tian Shan Sakas is significant when compared to Central Sakas; the Tagar display increased eastern hunter-gatherer (EHG) ancestry compared to all other Scythians. Lastly, the high genetic differentiation between western and eastern Scythians is emphasized by observing higher fixation index (F_{ST}) values between Hungarian Scythians and all Inner Asian Sakas (F_{ST} ranges from 0.24

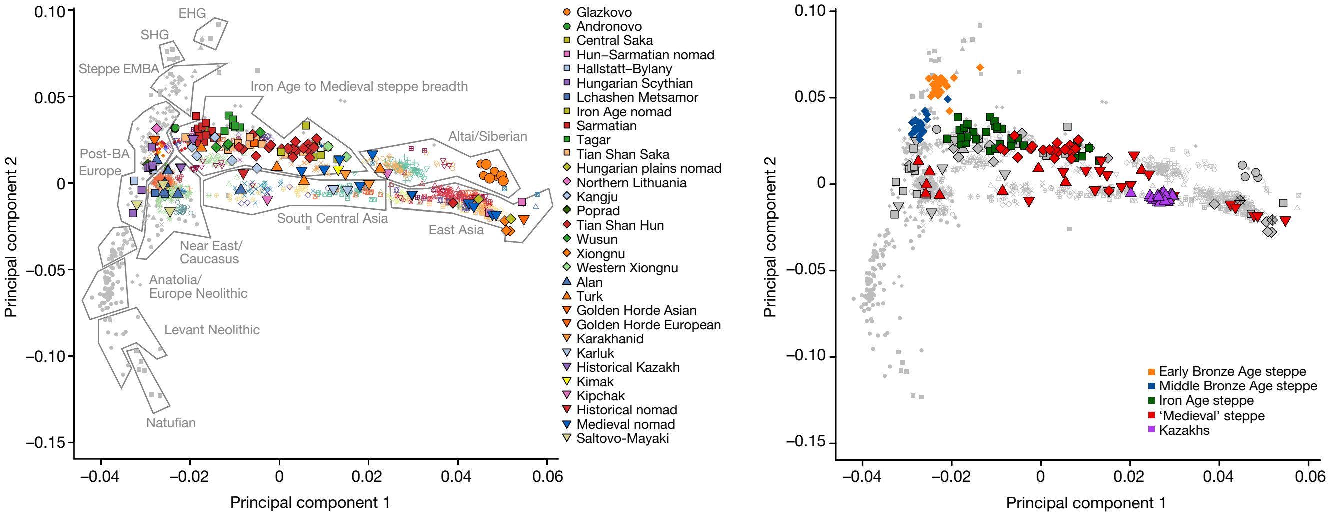


Fig. 2 | Principal component analyses. The principal components 1 and 2 were plotted for the ancient data analysed with the present-day data (no projection bias) using 502 individuals at 242,406 autosomal SNP positions. Dimension 1 explains 3% of the variance and represents a gradient stretching from Europe to East Asia. Dimension 2 explains 0.6% of the variance, and is a gradient mainly represented by ancient DNA starting from a 'basal-rich' cluster of Natufian hunter-gatherers and ending with EHGs. BA, Bronze Age; EMBA, Early-to-Middle Bronze Age; SHG, Scandinavian hunter-gatherers.

to 0.3) than observed among the different Inner Asian Sakas groups (F_{ST} ranges from 0.15 to 0.2) (Supplementary Table 6).

The qpAdm modelling[1] 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 southern Siberian hunter-gatherers, with almost equal proportions of Bronze Age herder (56%) and southern Siberian hunter-gatherer ancestry (44%). The southern Siberian Tagar show unequal ancestry contributions from Bronze Age herders (83.5%) and southern Siberian hunter-gatherers (7.5%), as well as an additional contribution of Mal'ta (MA1 individual)-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 southern Siberian hunter-gatherer ancestry (25%), and also 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[11], but instead indicate admixture between populations of Late Bronze Age herder descent and various local groups, consistent with the multiple origins model (model 3 described above).

Our data show that the culturally similar Scythians represented genetically structured groups within the Eurasian steppes. 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 who 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 of present-day eastern Turkmenistan.

The Xiongnu and the Hunnic expansions

Turkic language elements arguably first emerged among the Xiongnu nomads[15], a confederation of several nomadic tribes who occupied the eastern steppe from the third century BC. They are believed to be of East Asian ancestry[16,17], although ancient Y-chromosomal data have indicated a possibly heterogeneous population admixed with central steppe nomads[18]. Huns (third–fifth century AD) have previously been argued to derive directly from the Xiongnu[19], although others have claimed that there is no evidence connecting the two groups[20]. 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 affected 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. It has tentatively been suggested on the basis of the archaeological record that they belonged to the Iranian-speaking branch of the Indo-European language family[21].

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 (Fig. 2 and Extended Data Figs. 1, 5; in Fig. 2 these are labelled 'Xiongnu' and 'western Xiongnu', respectively). We find that Central Sakas are accepted as a source for these 'western-admixed' Xiongnu in a single-wave model. Consistent 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 cluster closely together in the PCA (Fig. 2).

We used *D*-statistics (Supplementary Information 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 compared to the Xiongnu (Extended Data Fig. 6). We tested for patterns of shared drift between the Xiongnu and the Wusun, the preceding Sakas and the slightly later Huns (second century AD). 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 Data Fig. 7). Similar results are seen with the contemporaneous and later Kangju groups that—as did 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 Table 7), together with Late Bronze Age pastoralists and the southern Siberian hunter-gatherers. We therefore suspect that the Wusun and Kangju groups are descendants of Bronze Age pastoralists that interacted with the civilization of the Bactria–Margiana archaeological complex in southern Uzbekistan and 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

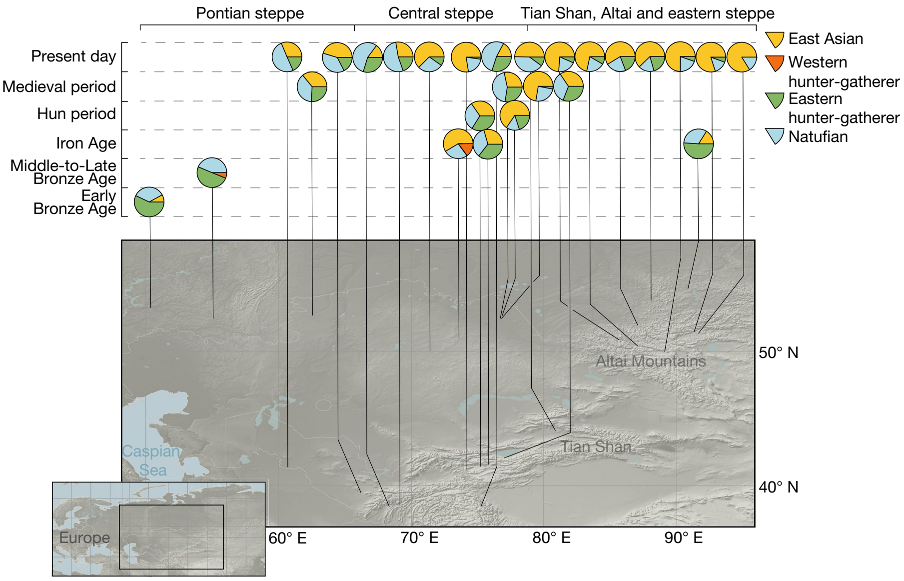


Fig. 3 | QpAdm results depicting the changes in ancestry across time in Central Asia. The changes reflect a gradual increase in East Asian ancestry in the central steppe nomads coupled to a decrease in ancestry associated with EHGs, starting at a high level in Yamnaya and finishing at a low level in present-day Kazakh and Kyrgyz individuals. The set of outgroups used is: Mbuti, Ust'Ishim, Clovis, Kostenki14 and Scandinavian hunter-gatherers.

invaded (see Supplementary Information section 3.6 for sex-biased admixture rates). As such our results support the contention that the disappearance of the Inner Asian Scythians and Sakas around two thousand years ago 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—that had remained on the south-eastern side of the Tian Shan mountains.

Repeated conquests and waves of East Asian impact

In the sixth century AD, the Hunnic Empire had been broken up and dispersed as the Turkic Khaganate assumed the military and political domination of the steppes[22,23]. Khaganates were steppe nomad political organizations that varied in size and became dominant during this period; they can be contrasted to the previous stateless organizations of the Iron Age[24]. The Turkic Khaganate was eventually replaced by a number of short-lived steppe cultures[25]. These included the Kipchak and the Tungusic Kimak populations, which spread southwards towards the Tian Shan mountains and westward towards the Ural mountains to form the Kimak Khaganate of the central steppe during the eighth to eleventh centuries AD[26]. During the eleventh century, the Kimak Khaganate was overthrown by local Kipchak groups, who in turn allied themselves with the Cuman of West Eurasia. Eventually the short-lived khaganates were overtaken by the Mongol Empire, which emerged through the unification of East Mongolian and Transbaikalian tribes and which expanded considerably during the rule of Genghis Khan in the thirteenth century AD[26,27] (Supplementary Information section 1).

We find evidence that elite soldiers associated with the Turkic Khaganate are genetically closer to East Asians than are the preceding Huns of the Tian Shan mountains (Supplementary Information section 3.7). We also find that one Turkic Khaganate-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 (Supplementary Information section 3.7). During the Kipchak period in the eleventh century AD, the domination of the central steppe was allegedly assumed by another group originating from the geographical area of Tuva. We present genomic data from two individuals from this period, one of whom shows increased East Asian ancestry, whereas the other has pronounced European ancestry (samples DA23 and DA179, respectively, in Supplementary Information section 4). These individuals date to the Cuman–Kipchak alliance, which incorporated both the western and eastern steppe. For the period in which the region became incorporated into the Karakhanid Khanate—which encompassed present-day regions of Uzbekistan, Tajikistan, Kazakhstan and Kyrgyzstan—*D*-statistics identify a small influx of East Asian ancestry compared to the earlier Turk period. Consistent with this, nomads in the Karakhanid period are shifted towards East Asians compared to earlier Turks in the PCA plot (Fig. 2 and Extended Data Fig. 8). Additionally, we analysed ten culturally unaffiliated Medieval-period nomads, most of whom showed pronounced East Asian ancestry, albeit in very different proportions (Extended Data Fig. 8). We also find the presence of an individual of West Eurasian descent buried together with members of Jochi Khan's Golden Horde army from the Ulytau mountains (see Supplementary Information 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 Turkic 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 (Fig. 2) separating Eastern and Western descent. On the basis of this notable heterogeneity, we suggest that during the Medieval period 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 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 steppes, large areas of Europe were depopulated owing to the Justinian plague pandemic[28]. Although 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

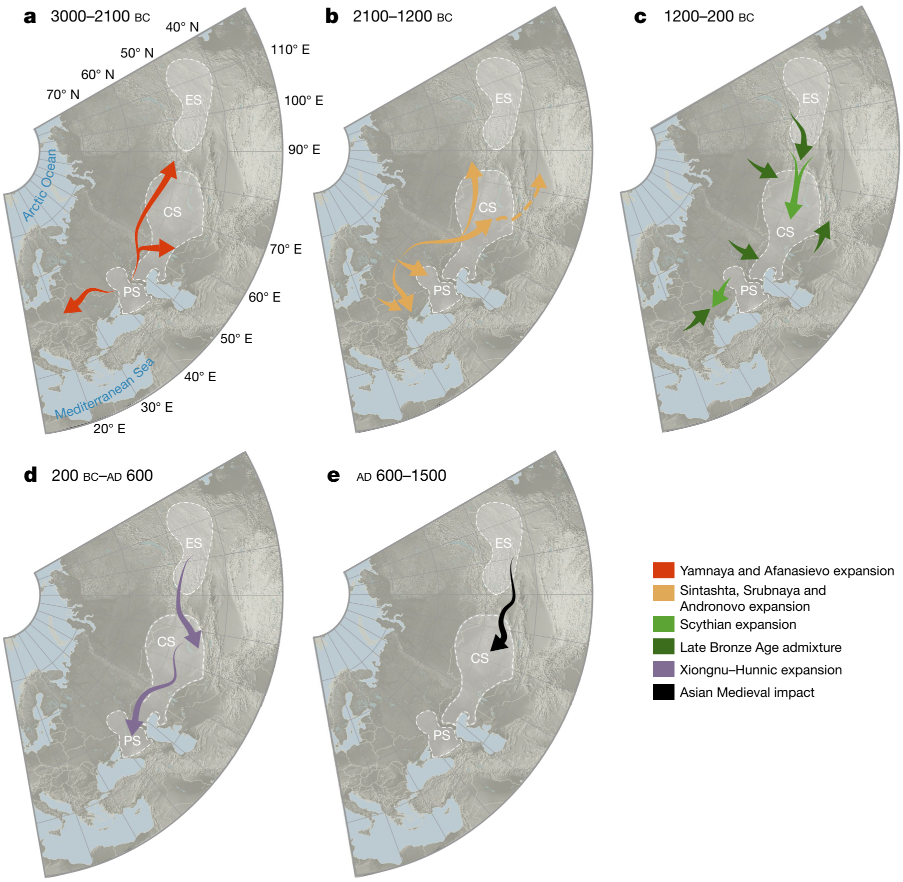


Fig. 4 | Summary map. Depictions of the five main migratory events associated with the genomic history of the steppe pastoralists from 3000 BC to the present. **a**, Depiction of Early Bronze Age migrations related to the expansion of Yamnaya and Afanasievo culture. **b**, Depiction of Late Bronze Age migrations related to the Sintashta and Andronovo horizons. **c**, Depiction of Iron Age migrations and sources of admixture. **d**, Depiction of Hun-period migrations and sources of admixture. **e**, Depiction of Medieval migrations across the steppes.

in Xinjiang in China, thus pointing to a possible Inner Asian origin of the Justinian plague.

We find that two individuals, DA101 and DA147 (see Supplementary Information section 7), show detectable levels of *Yersinia pestis* DNA, compatible with the characterization of the full genome sequence at 8.7× and 0.24× coverage. The first individual (DA101) is a Hun from the Tian Shan mountains and dates to approximately AD 180, and the second individual (DA147) is from the Alan culture from North Ossetia and is estimated archaeologically to date to the sixth–ninth century AD. The genome of 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 about AD 530[29] (Extended Data Fig. 9). As expected, the Tian Shan strain contained the *ymt* gene reported to be missing in the more-ancestral Bronze Age plague strains[31]. The strain also displayed the loss of function mutations in *pde2*, *pde3*, *rcsA* and *ureD* that are required for flea transmission in the traditional 'blocked flea' model[32] (Extended Data Fig. 9). This, coupled with a fully functional plasminogen activator gene, indicates that the 'Hunnic' plague strain had full bubonic capability and flea transmissibility.

The fact that we find a higher number of strain-specific variants in the Aschheim strain is consistent with the difference in sampling time (approximately AD 180 versus approximately AD 530) and the potentially multiple replication cycles associated with pandemics[30]. 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 owing to an increase in replication rate[30]. Therefore, our observation of an accelerated mutation rate is consistent with this hypothesis and supports the idea 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 in Qinghai[30] and the clade basal to the Justinian plague (0.ANT1) is from Xinjiang (China), two areas close to the Tian Shan mountains, 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 steppes.

Discussion

The overall population history that formed the genetic composition of present-day steppe populations is illustrated in Fig. 3, in which 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 Data Fig. 10 and Supplementary Information section 3.6). In Fig. 4, we present the main migratory patterns. Our findings fit well with current insights from the historical linguistics of this region (Supplementary Information section 2). The steppes were probably largely Iranian-speaking in the first and second millennia BC. This is supported by the split of the Indo-Iranian linguistic branch into Iranian and Indian[33], the distribution of the Iranian languages, and the preservation of Old Iranian loanwords in Tocharian[34]. The wide distribution of the Turkic languages from Northwest China, Mongolia and Siberia in the east to Turkey and Bulgaria in the west implies large-scale migrations out of the homeland in Mongolia since about 2,000 years ago[35]. The diversification within the Turkic languages suggests that several waves of migration occurred[36] and, on the basis of the effect of local languages, gradual assimilation to local populations had previously been assumed[37]. The East Asian migration starting with the Xiongnu accords well with the hypothesis that early Turkic was the major language of Xiongnu groups[38]. Further migrations of East Asians westwards find a good linguistic correlate in the influence of Mongolian on Turkic and Iranian in the last millennium[39]. As such, the genomic history of the Eurasian steppes is the story of a gradual transition from Bronze Age pastoralists of West 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 number PRJEB20658 (ERP022829). Single nucleotide polymorphism data for present-day populations are available, after ethical validation, from the European Genome-Phenome Archive (EGA, https://www.ebi.ac.uk/ega/) under accession number EGAS00001002926. Plague reads were deposited in the European Nucleotide Archive (ENA) under accession number PRJEB25891.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0094-2.

Received: 18 April 2017; Accepted: 3 April 2018;
Published online 9 May 2018.

1. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
2. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
3. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
4. Chlenova, N. L. in *The Archaeology of the Steppes: Methods and Strategies* (ed. Genito, B.) 499–540 (Istituto Universitario Orientale, Naples, 1994).
5. Grakov, B. N., Yelagina, N. G. & Yatsenko, I. V. *The Early Iron Age* (Moscow State Univ. Press, Moscow, 1977).
6. Kristiansen, K. *Europe Before History* (Cambridge Univ. Press, Cambridge, 2000).
7. Parzinger, H. *Die Frühen Völker Eurasiens: vom Neolithikum bis zum Mittelalter* (CH Beck, München, 2006).
8. Alekseev, A. in *The Golden Deer of Eurasia: Scythian and Sarmatian Treasures from the Russian Steppes* (eds Aruz, J. et al.) 41–47 (The Metropolitan Museum of Art, New York, 2006).
9. Yablonsky, L. in *The Golden Deer of Eurasia: Scythian and Sarmatian Treasures from the Russian Steppes* (eds Aruz, J. et al.) 24–31 (The Metropolitan Museum of Art, New York, 2006).
10. Bashilov, V. A. & Yablonsky, L. T. in *Kurgans, Ritual Sites, and Settlements: Eurasian Bronze and Iron Age* (eds Davis-Kimball, J. et al.) 9–12 (Archaeopress, Oxford, 2000).
11. Unterländer, M. et al. Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat. Commun.* **8**, 14615 (2017).
12. Frachetti, M. D. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr. Anthropol.* **53**, 2–38 (2012).
13. Kohl, P. L. Shared social fields: evolutionary convergence in prehistory and contemporary practice. *Am. Anthropol.* **110**, 495–506 (2008).
14. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
15. Dybo, A. V. *Lingvističeskie kontakty rannix tjurkov. Leksičeskij fond* (Vostočnaja Literatura, Moscow, 2007).
16. Keyser-Tracqui, C., Crubézy, E. & Ludes, B. Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *Am. J. Hum. Genet.* **73**, 247–260 (2003).
17. Keyser-Tracqui, C., Crubézy, E., Pamzsav, H., Varga, T. & Ludes, B. Population origins in Mongolia: genetic structure analysis of ancient and modern DNA. *Am. J. Phys. Anthropol.* **131**, 272–281 (2006).
18. Kim, K. et al. A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast Mongolia. *Am. J. Phys. Anthropol.* **142**, 429–440 (2010).
19. Pohl, W. in *A Companion to Ethnicity in the Ancient Mediterranean* (ed. McInerney, J.) 555–568 (Wiley Blackwell, Chichester, 2014).
20. De la Vaissière, É. Huns et Xiongnu. *Cent. Asiat. J.* **49**, 3–26 (2005).
21. Mallory, J. P. *In Search of the Indo-Europeans. Language, Archaeology and Myth* (Thames & Hudson, London, 1989).
22. Sinor, D. in *The Cambridge History of Early Inner Asia* (ed. Sinor, D.) 285–316 (Cambridge Univ. Press, Cambridge, 1990).
23. Findley, C. V. *The Turks in World History* (Oxford Univ. Press, Oxford, 2004).
24. Kradin, N. in *Xiongnu Archaeology: Multidisciplinary Perspectives of the First Steppe Empire in Inner Asia* (eds Brosseder, U. & Miller, B. K.) 77–96 (Universität Bonn, Bonn, 2011).
25. Golden, P. B. *An Introduction to the History of the Turkic Peoples: Ethnogenesis and State-Formation in Medieval and Early Modern Eurasia and the Middle East* (Harrassowitz, Wiesbaden, 1992).
26. Hildinger, E. *Warriors of the Steppe: a Military History of Central Asia, 500 BC to 1700 AD* (Da Capo Press, Cambridge, 1997).
27. Kradin, N. N. & Skrynnikova, T. D. *Imperiya Imperija Chingis Čingis-Khana Xana [The Genghis Khan Empire]* (Vostočnaja Literatura, Moscow, 2006).
28. Little, L. K. *Plague and the End of Antiquity: the Pandemic of 541–750* (Cambridge Univ. Press, Cambridge, 2007).
29. Wagner, D. M. et al. *Yersinia pestis* and the plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
30. Cui, Y. et al. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis. Proc. Natl Acad. Sci. USA* **110**, 577–582 (2013).
31. Rasmussen, S. et al. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
32. Sun, Y.-C. C., Jarrett, C. O., Bosio, C. F. & Hinnebusch, B. J. Retracing the evolutionary path that led to flea-borne transmission of *Yersinia pestis. Cell Host Microbe* **15**, 571–582 (2015).
33. Kuz'mina, E. E. *The Origin of the Indo-Iranians* (Brill, Leiden, 2007).
34. Tremblay, X. Irano-Tocharica et Tocharo-Iranica. *Bull. Sch. Orient. Afr. Stud.* **68**, 421–449 (2005).
35. Nichols, J. in *Language Contact in Times of Globalization* (eds Hasselblatt, C. et al.) 177–195 (Rodopi, Amsterdam, 2011).
36. Johanson, L. in *The Turkic Languages* (eds Johanson, L. & Csató, É. Á.) 81–125 (Routledge, London, 1998).
37. Johanson, L. in *The Handbook of Language Contact* (ed. Hickey, R.) 652–672 (Wiley-Blackwell, Chichester, 2010).
38. Janhunen, J. *Manchuria: an Ethnic History* (The Finno-Ugrian Society, Helsinki, 1996).
39. Doerfer, G. *Türkische und Mongolische Elemente im Neupersischen 1–4* (Harrassowitz, Wiesbaden, 1963–1975).
40. Goldberg, A. et al. Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. *Proc. Natl Acad. Sci. USA* **114**, 2657–2662 (2017).

Acknowledgements We thank K. Magnussen, L. Petersen, C. Mortensen and A. Seguin-Orlando at the Danish National Sequencing Centre for producing the analysed sequences; P. Reimer and S. Hoper at the 14Chrono Center Belfast for providing accelerator mass spectrometry dating; S. Hackenbeck for discussing palaeodietary reconstructions; D. Christiansen Appelt, B. Heyerdahl, the Explico Foundation team, J. Isakova, B. Daulet, A. Tairov, N. Abduov, B. Tudiyarov, V. Volkov, M. Akchurin, I. Baimukhan, N. Namdakov, Y. Yusupov, E. Ramankulov, A. Nurgaziyev and A. Kusaev for important assistance in fieldwork; J. Stenderup, P. V. Olsen and T. Brand for technical assistance in the laboratory; all involved archaeologists, historians and geographers from Kazakhstan: A. Suslov, I. Erofeeva, E. Nurmaganbetov, B. Kozhakhmetov, N. Loman, Y. Parshin, S. Ladunskiy, M. Bedelbaeva, A. Marcsik, O. Gábor, M. Půlpán, Y. Kubeev, R. Zhumashev, K. Omarov, S. Kasymov and U. Akimbayeva; P. Rodzianko for creating the initial contact between P.d.B.D., S.E. and E.U.; and S. Jacobsen and J. O'Brien for translating and proofreading Russian contributions. E.W. thanks St. John's College, Cambridge for support and for providing an environment facilitating scientific discussions. B.Boldg. thanks the Taylor Family-Asia Foundation Endowed Chair in Ecology and Conservation Biology. The project was funded by the Danish National Research Foundation (E.W.), the Lundbeck Foundation (E.W.) and KU2016 (E.W.).

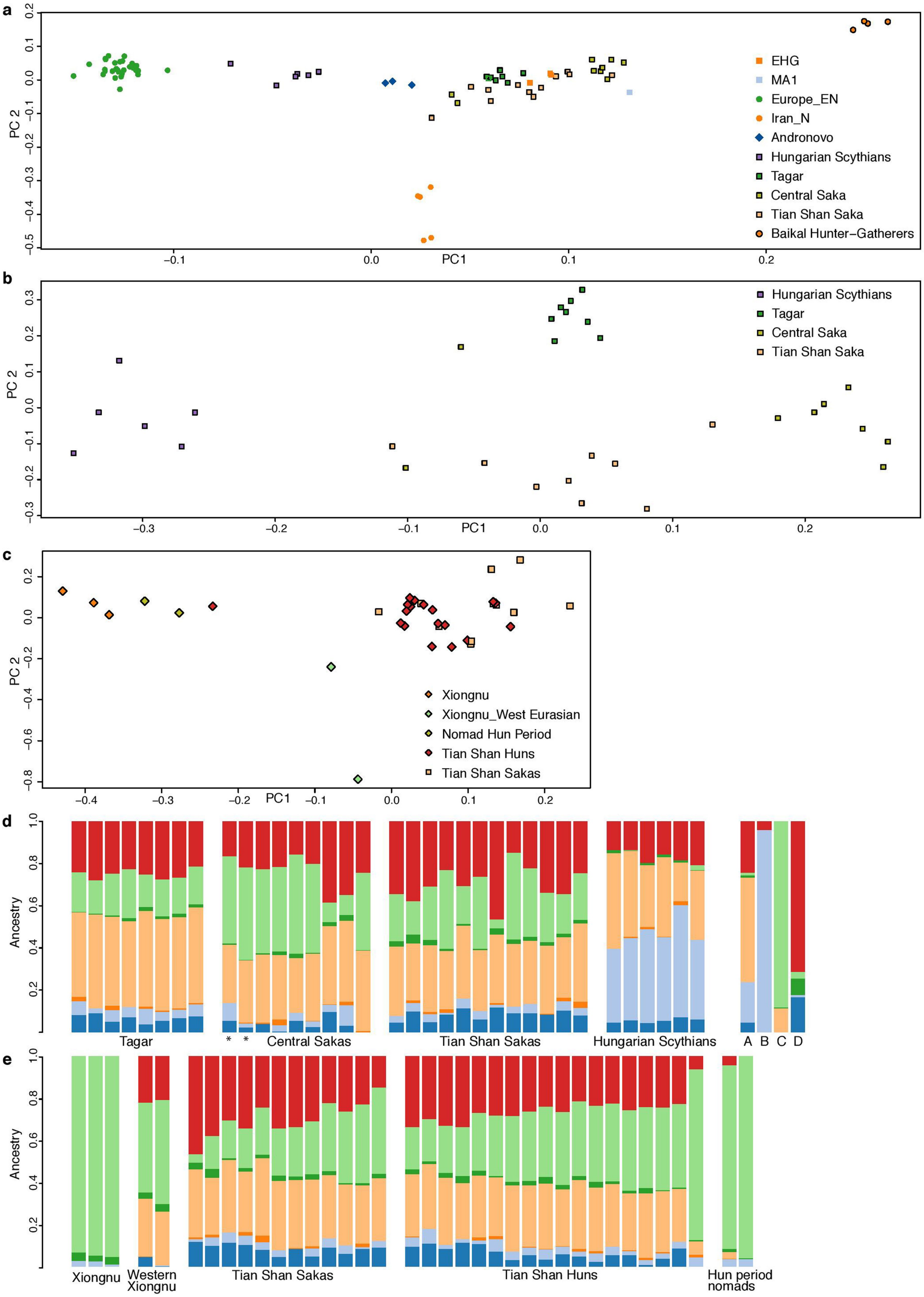
Reviewer information *Nature* thanks T. Higham, D. Anthony, B. Shapiro, R. Dennell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions E.W. initiated and led the study. P.d.B.D., E.W., E.U. and E.H. designed the study. P.d.B.D. and N.M. produced the data. P.d.B.D., N.M., S.R., M.S., G.R., T.Ko., A.Gol., M.W.P., A.G.P. and K.N. analysed or assisted in analysis of data. A.M.T. and M.A.W.S. provided an overview of major Y-chromosomal haplogroups in Supplementary Information Section 8. P.d.B.D., E.W. and K.K. interpreted results with considerable input from M.S., R.N., M.P., N.K., S.R., L.O., M.E.A. and J.V.M.-M. P.d.B.D., E.W., K.K., M.P. and S.R. wrote the manuscript with considerable input from N.K., L.H., M.S., R.N., M.E.A., L.O. and J.V.M.-M., with contributions from all authors. P.d.B.D., M.E.A., L.O., E.U., N.B., V.L., G.A., K.A., A.Ald., A.Alp., G.B., V.I.B., A.B., B.Boldb., B.Boldg., C.D., S.E., D.E., R.D., E.D., V.E., K.M.F., A.Gor., A.Gr., H.H., T.H., Z.K., R.K., E.K., A.Ko., T.Ku., A.Ku., I.K., N.L., A.M., V.K.M., I.V.M., I.M., E.M., V.M., G.M., B.N., Z.O., I.P., K.P., V.S., I.S., A.L., K.-G.S., T.S., K.T., A.T., T.T., D.V., L.Y., S.U., V.V., A.W. and E.H. excavated, curated, sampled and/or described analysed skeletons; all authors contributed to final interpretation of data.

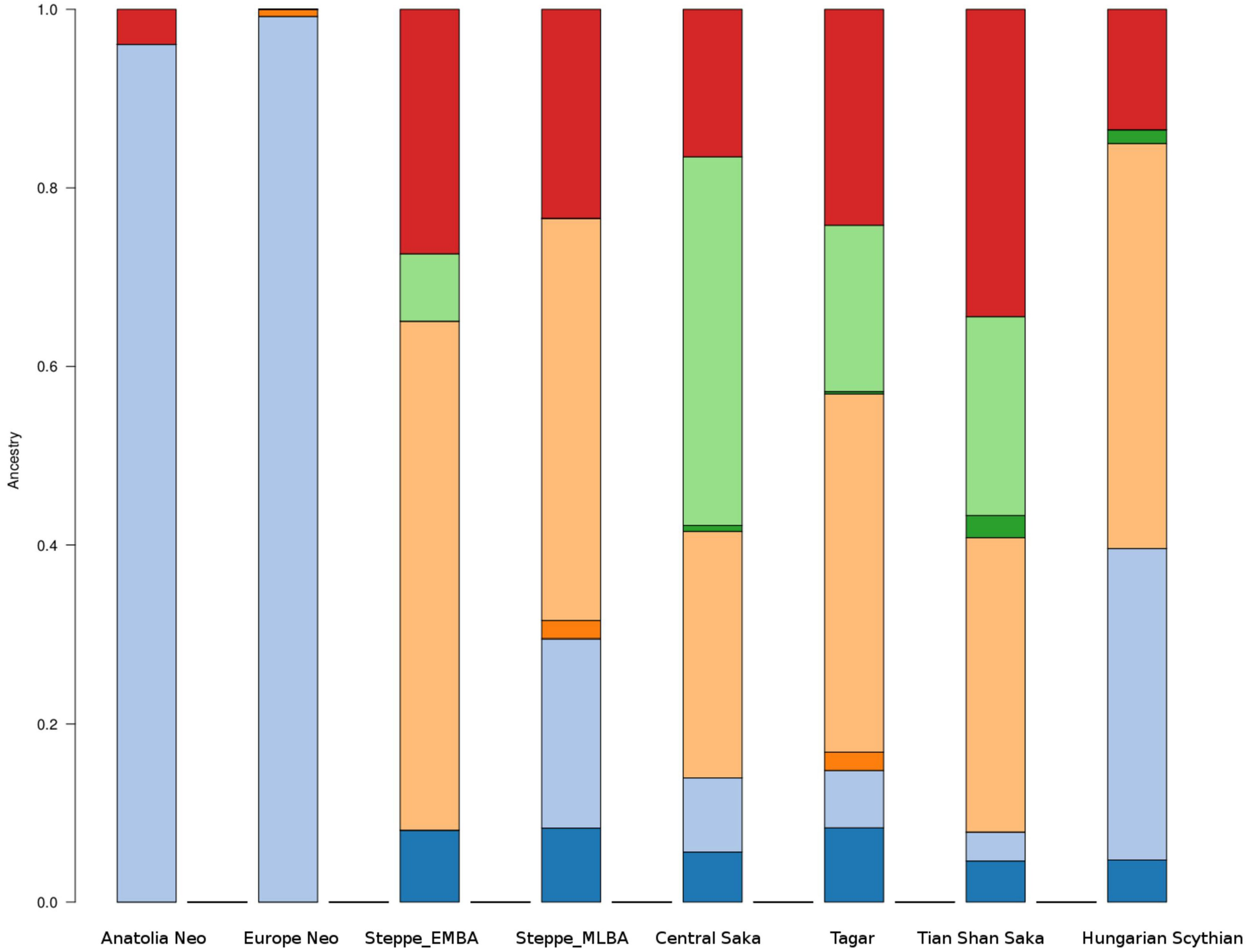
Competing interests The authors declare no competing interests.

Additional information

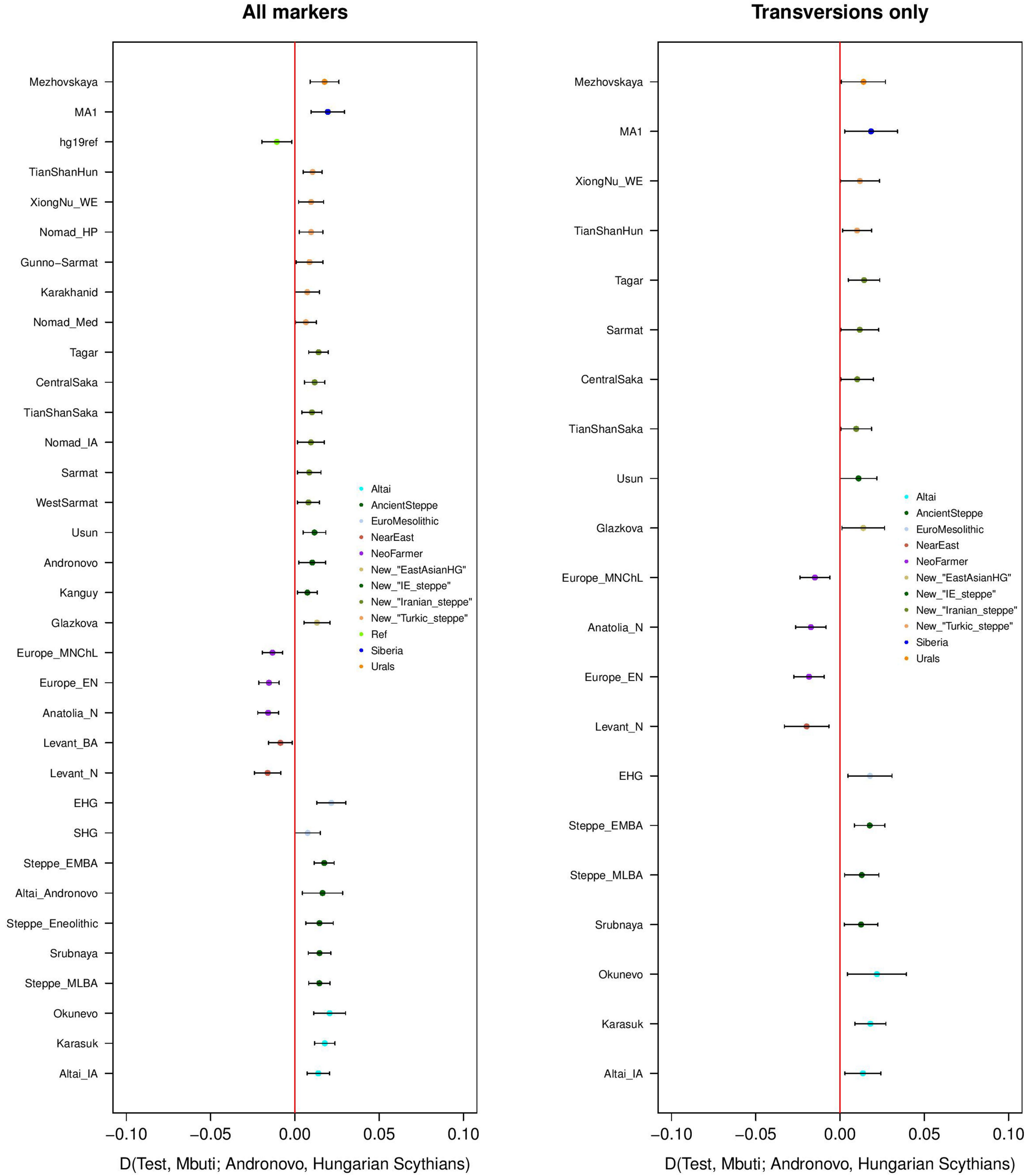
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0094-2.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0094-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to E.W.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



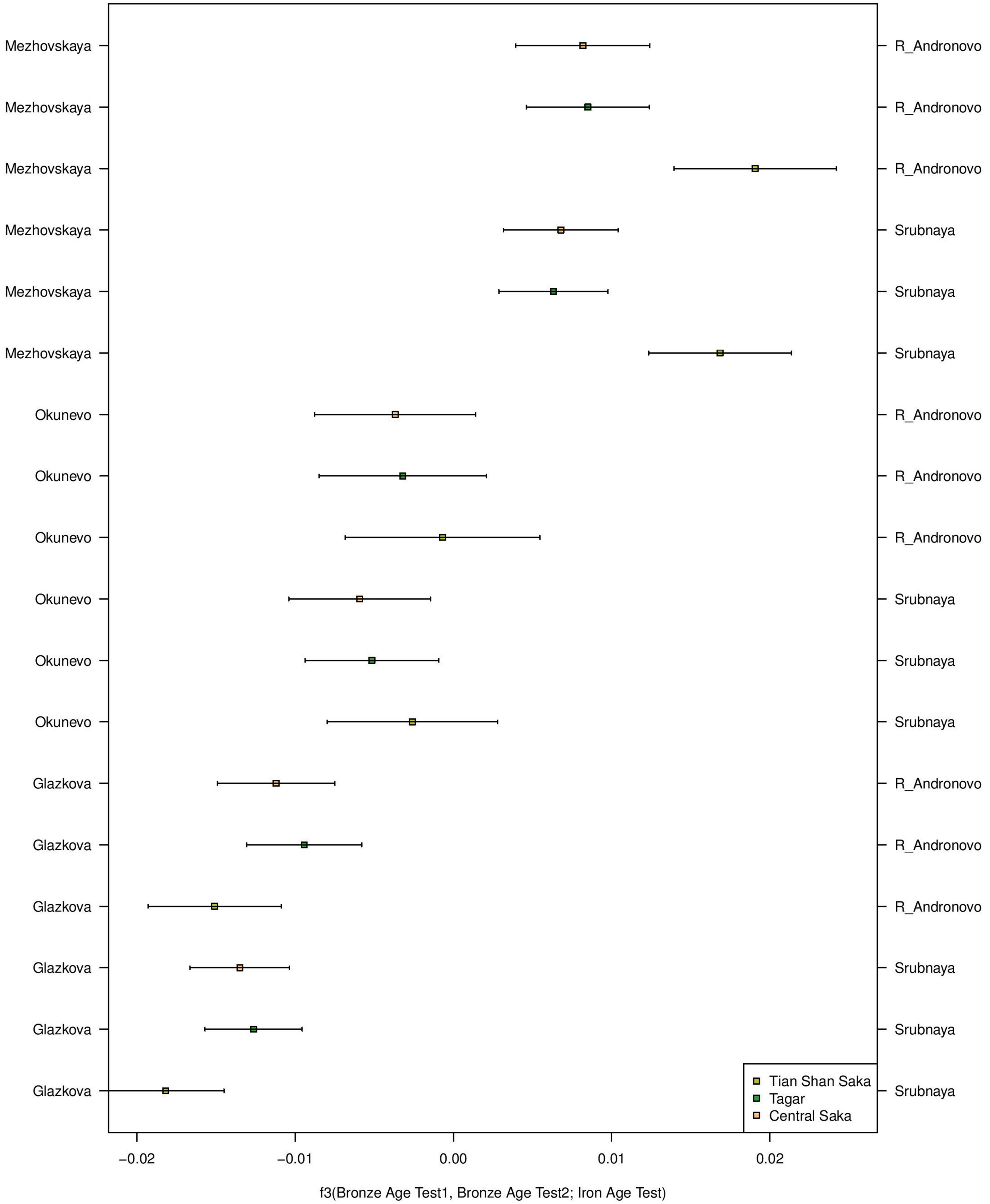
Extended Data Fig. 1 | Analyses of Iron Age clusters. a, PCA of Iron Age nomads and ancestral sources, explaining the diversity between them using 74 individuals at 242,406 autosomal single nucleotide polymorphism (SNP) positions. **b**, PCA of Iron Age nomads alone using 29 individuals at 242,406 autosomal SNP positions. **c**, PCA of Xiongnu, ‘Western’ Xiongnu, Tian Shan Huns, Nomads Hun Period, and Tian Shan Sakas, using 39 individuals at 242,406 autosomal SNP positions. **d**, Model-based clustering at $K = 7$ illustrating differences in ancestral proportions. Labelled individuals: A, Andronovo; B, Neolithic European (Europe_EN, in **a**); C, Baikal hunter-gatherers; D, Neolithic Iranian (Iran_N, in **a**). 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 asterisk indicates an individual flagged as a genetic outlier. **d**, **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). Panel **d** is focused on the Iron Age, while **e** is focused on the transition to the Hun period.



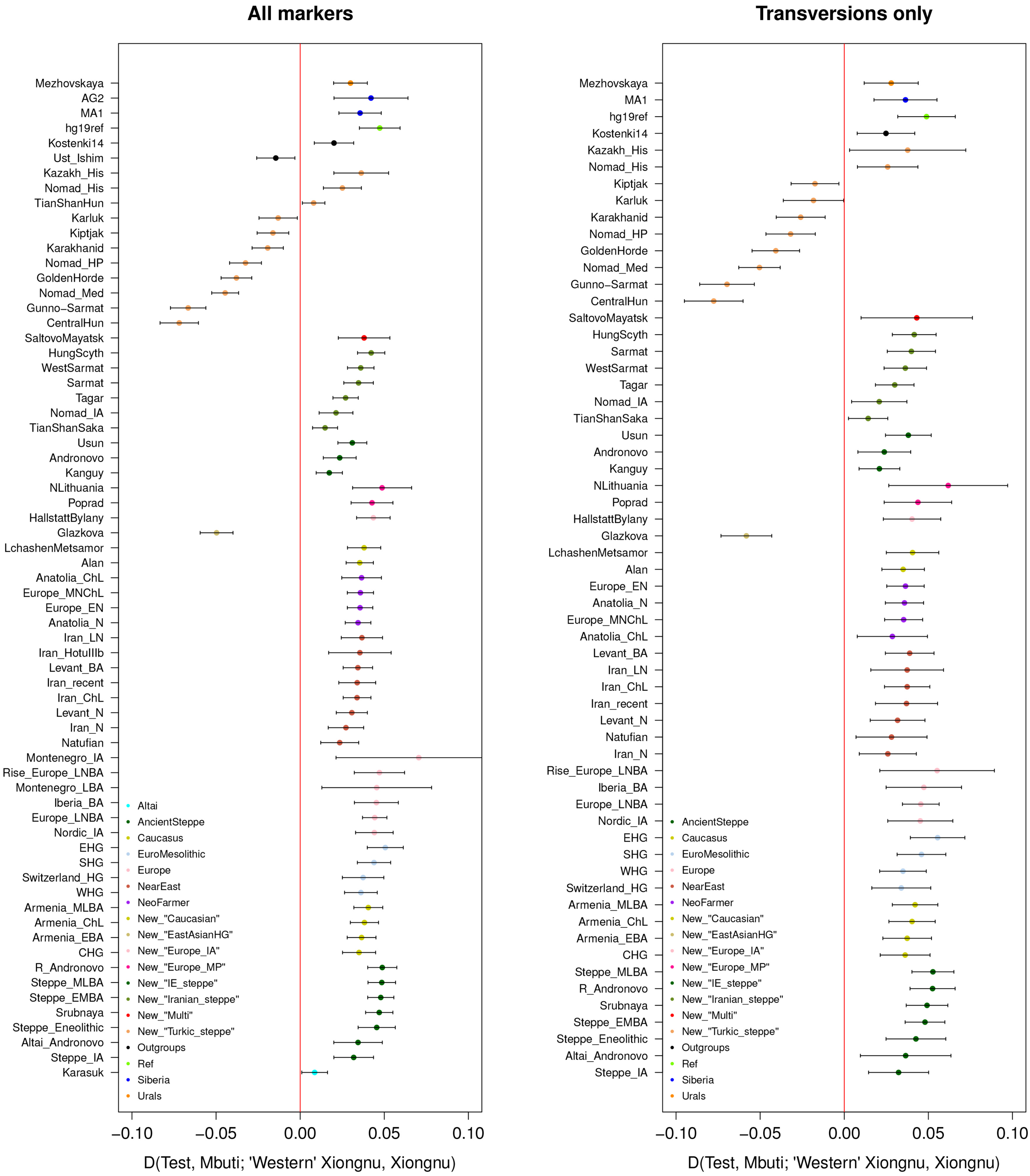
Extended Data Fig. 2 | Illustration of shared ancestry between Neolithic farmers and Iron Age nomads. Results for model-based clustering analysis at $K = 7$, plotting only one individual from relevant groups, to illustrate shared ancestry between Neolithic farmers from Europe, Late Bronze Age nomads and Iron Age nomads, not shared with Early Bronze Age nomads. MBLA, Middle-to-Late Bronze Age; Neo, Neolithic.



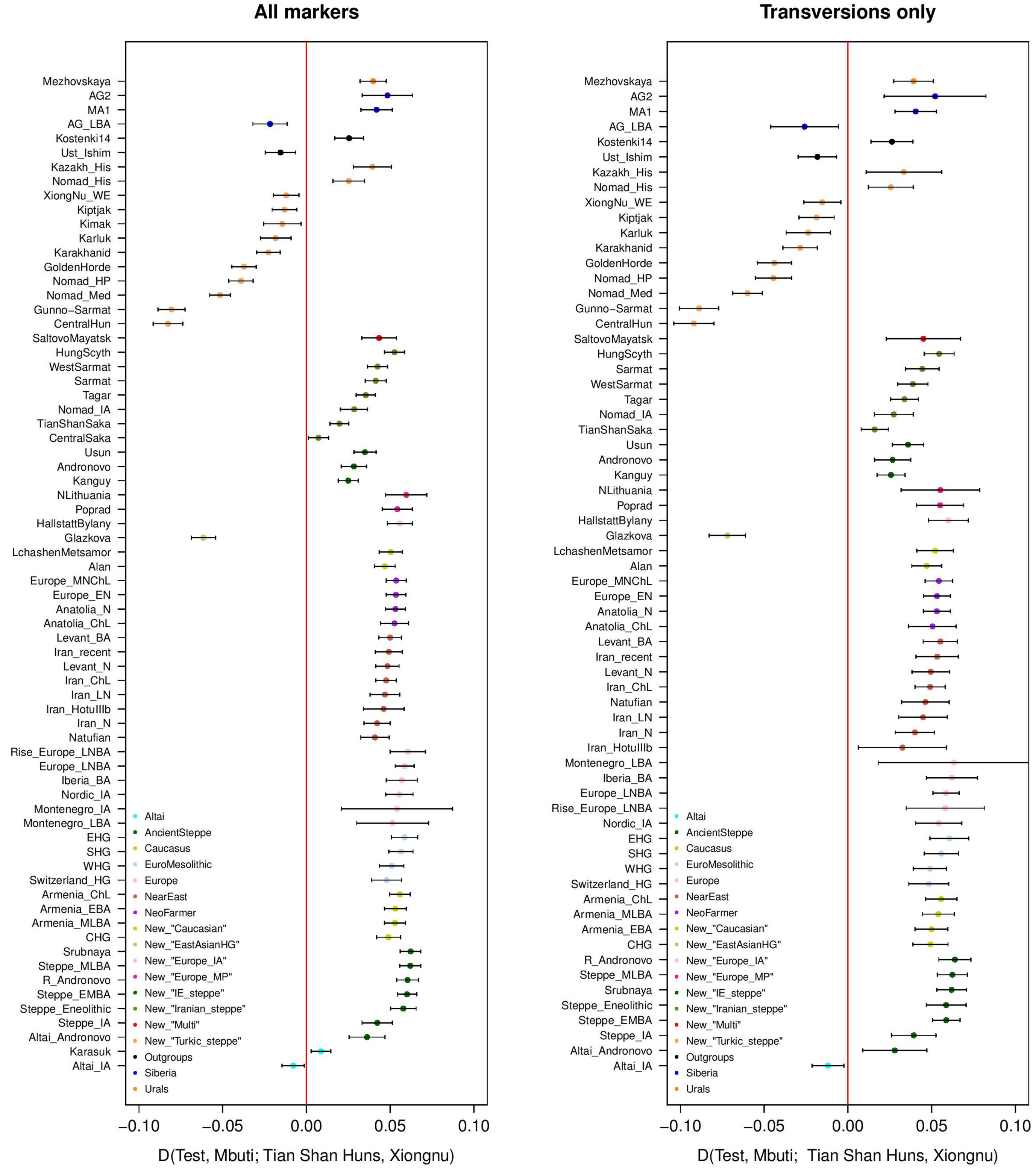
Extended Data Fig. 3 | Illustration of gene flow into Hungarian Scythians. We represent all *D*(Test, Mbuti; Andronovo, Hungarian Scythians) that deviate significantly from 0 (that is, higher than 3× the standard errors). The reported numbers are the *D*-statistics and the 3 standard errors were plotted as error bars. The number of individuals per population can be found in Supplementary Tables 3, 4.



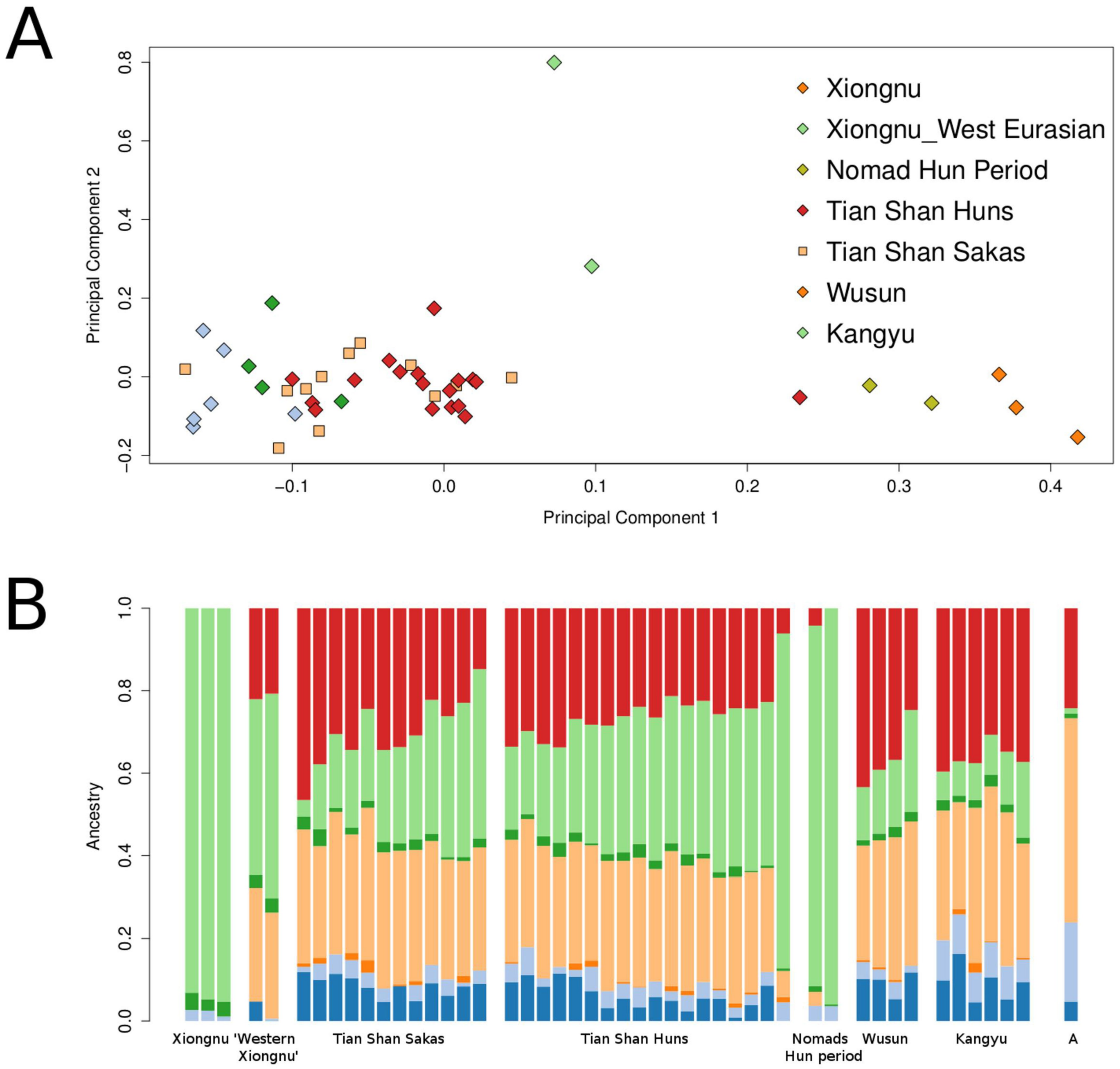
Extended Data Fig. 4 | Illustration of negative admixture f_3 statistics for Iron Age populations. Plot shows $f3$(Bronze Age Test 1, Bronze Age Test 2; Iron Age Test). The reported numbers are of the f_3 statistics, and the 3 standard errors were plotted as errors bars. The number of individuals per population can be found in Supplementary Table 3.



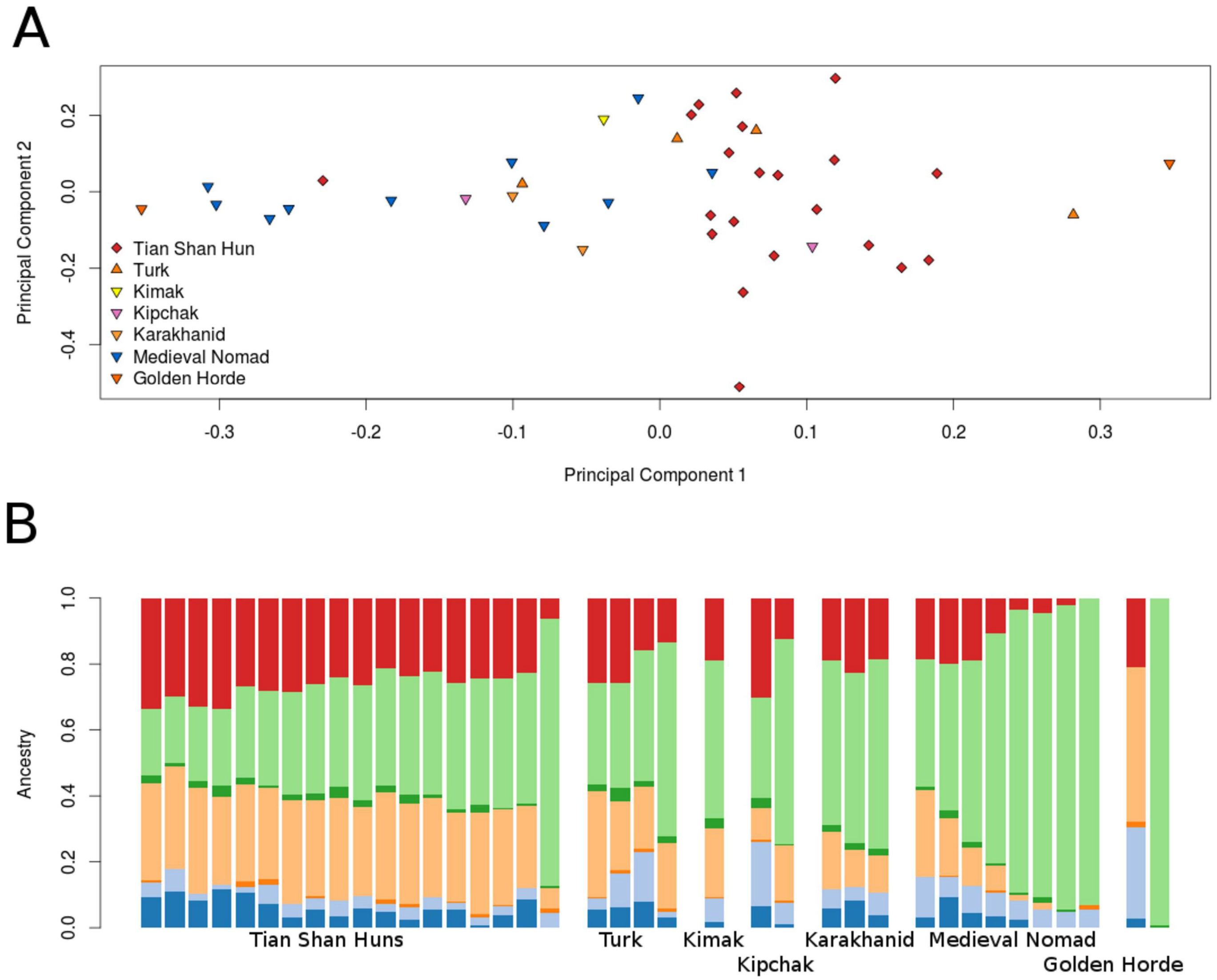
Extended Data Fig. 5 | Illustration of West Eurasian gene flow into groups forming the Xiongnu culture. We represent all *D*(Test, Mbuti; 'Western' Xiongnu, Xiongnu) that deviate significantly from 0 (that is, higher than 3× the standard errors). The reported numbers are the *D*-statistics and the 3 standard errors were plotted as error bars. The number of individuals per population can be found in Supplementary Tables 3, 4.



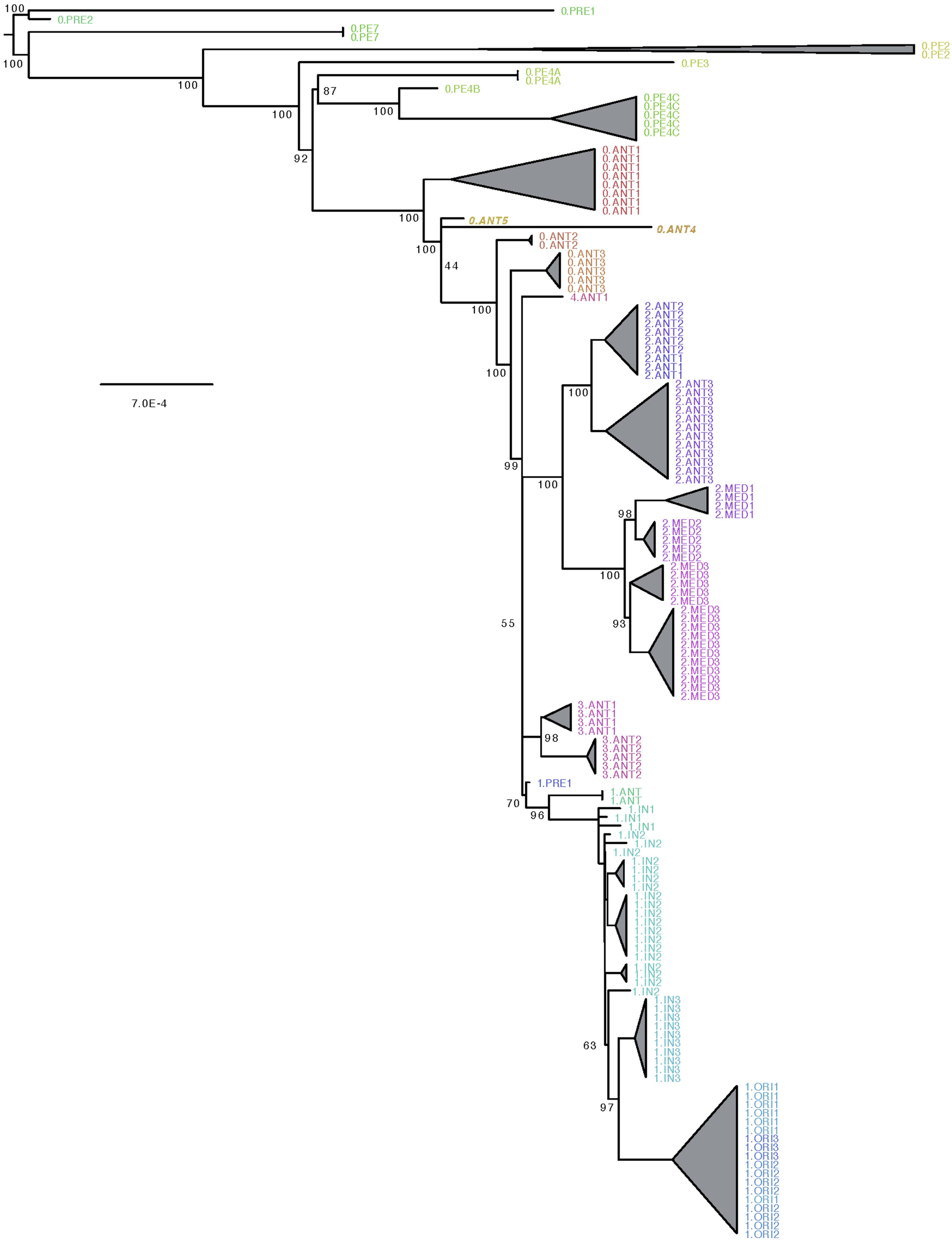
Extended Data Fig. 6 | Illustration of West Eurasian ancestry in early Tian Shan Huns. We represent all *D*(Test, Mbuti; Tian Shan Huns, Xiongnu) that deviate significantly from 0 (that is, higher than 3× the standard errors). The reported numbers are the *D*-statistics and the 3 standard errors were plotted as error bars. The number of individuals per population can be found in Supplementary Tables 3, 4.



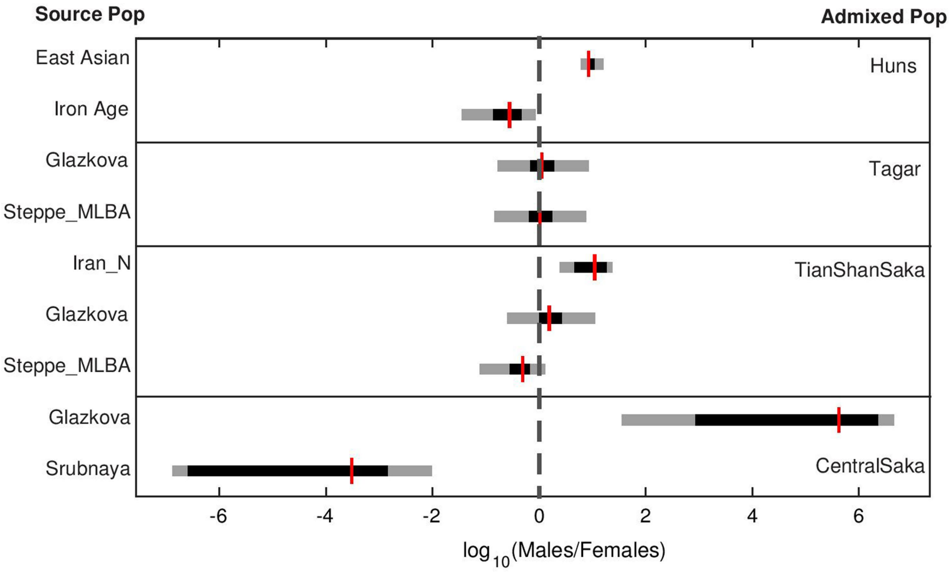
Extended Data Fig. 7 | Analyses of Xiongnu and Hun period population clusters. a, PCA of Xiongnu, 'Western' Xiongnu, Tian Shan Huns, Hun-period nomads, Tian Shan Sakas, Kangju and Wusun, including 49 individuals analysed at 242,406 autosomal SNP positions. **b**, Results for model-based clustering analysis 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). Individual A is a southern Siberian individual associated with the Andronovo culture.



Extended Data Fig. 8 | Analyses of Turk- and Medieval-period population clusters. a, PCA of Tian Shan Hun, Turk, Kimak, Kipchack, Karakhanid and Golden Horde, including 28 individuals analysed 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 Data Fig. 9 | Maximum likelihood phylogenetic reconstruction of *Y. pestis*. This tree reveals the basal position of the Tian Shan sample (0.ANT5, DA101, AD 186) compared to the Justinian plague sample (0.ANT4, A120, AD 536). These two samples are shown in orange italics. Other ancient plague samples included in the tree are Bronze Age samples (0.PRE1 and 0.PRE2) and a Black Death sample (1.PRE1). Numbers on nodes indicate bootstrap support (not all of which are shown, for clarity) and certain branches have been collapsed for clarity. Branch lengths are substitutions per site.



Extended Data Fig. 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 a previously published method[40]. 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 logarithms 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.

nature research

Corresponding author(s): Eske Willerslev Nature 2017-04-05008C

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No specific software was used for data collection. All software used in this study is listed below.
Data analysis	All software used in this work is publicly available. Corresponding publications are cited in the main text and supplementary material. List of software and respective versions: CASAVA v1.8.2 AdapterRemoval v2.1.3 bwa v0.7.10 bwa mem 0.7.10 picard tools v1.127 bamUtil v1.0.14 samtools v1.3.1 GATK v3.3.0 and v 3.6* pysam 0.7.4 (python module) bedtools 2.27.1 mapDamage2.0 contamMix v1.0-5 SHRiMP 2.2.3 YFitter v0.2 Haplogrep 2.0

ANGSD v0.915
PRANK v.150803
BEASTv1.8.2
schmutzi
admixtools v4.1
NGSrelate
plink v1.07* and v1.9
ADMIXTURE v1.3
RAxML-8.1.15
SnpEff
SPAdes-3.9.0
R 3.2.3
python 2.7.12
perl v5.22.1
CALIB
NOTE: Versions marked with * were used for Y chromosome analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not rely on statistical methods to predetermine sample sizes. Sample sizes in ancient population genetic studies are limited by the number of samples yielding endogenous DNA proportions amenable to whole genome sequencing.
Data exclusions	We selected 137 samples for whole-genome sequencing, out of all screened samples, based on their endogenous content and low contamination estimates. These criteria are described in detail in Supplementary Section 3.1. Furthermore, closely related individuals were excluded from analyses requiring population allele frequencies.
Replication	We did not attempt to specifically replicate experimental findings. But we note that samples from the same population carry similar genetic signatures. Moreover, genome-wide data allows for the analysis of multiple realisations of the sample history, by studying hundreds of thousands of SNP sites.
Randomization	No experimental groups or effect sizes were measured in this study, thus we did not implement any random group assignment.
Blinding	No blinding techniques were implemented, as exprimental group assignment is not relevant for popluation genetic history studies of this kind.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☐	☒ Human research participants

Human research participants

Policy information about studies involving human research participants

Population characteristics	No experimental procedures were carried out on human participants. We genotyped 502 individuals from 16 self-reported ethnicities from Altai, Central Asia, Siberia and the Caucasus. Sampling procedures are detailed in Supplementary Section 5.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging