www.sciencemag.org/cgi/content/full/science.aar7711/DC1

Supplementary Materials for

The first horse herders and the impact of early Bronze Age steppe expansions into Asia

Peter de Barros Damgaard*, Rui Martiniano*, Jack Kamm*, J. Víctor Moreno-Mayar*, Guus Kroonen, Michaël Peyrot, Gojko Barjamovic, Simon Rasmussen, Claus Zacho, Nurbol Baimukhanov, Victor Zaibert, Victor Merz, Arjun Biddanda, Ilja Merz, Valeriy Loman, Valeriy Evdokimov, Emma Usmanova, Brian Hemphill, Andaine Seguin-Orlando, Fulya Eylem Yediay, Inam Ullah, Karl-Göran Sjögren, Katrine Højholt Iversen, Jeremy Choin, Constanza de la Fuente, Melissa Ilardo, Hannes Schroeder, Vyacheslav Moiseyev, Andrey Gromov, Andrei Polyakov, Sachihiro Omura, Süleyman Yücel Senyurt, Habib Ahmad, Catriona McKenzie, Ashot Margaryan, Abdul Hameed, Abdul Samad, Nazish Gul, Muhammad Hassan Khokhar, O. I. Goriunova, Vladimir I. Bazaliiskii, John Novembre, Andrzej W. Weber, Ludovic Orlando, Morten E. Allentoft, Rasmus Nielsen, Kristian Kristiansen, Martin Sikora, Alan K. Outram, Richard Durbin†, Eske Willerslev†

*These authors contributed equally to this work.
†Corresponding author. Email: rd109@cam.ac.uk (R.D.); ewillerslev@snm.ku.dk (E.W.)

Published 9 May 2018 on *Science* First Release
DOI: 10.1126/science.aar7711

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aar7711/DC1)

Tables S3 and S14 as Excel files

Table of Contents

Supplementary Text

S1: Sample description

S1.1 Skeletal materials from Botai

Recent studies focusing on the archaeology of Copper Age Botai culture (~3500–3000 BCE) provide strong evidence for the practice of horse domestication. First, examination of dental pathologies in Botai horses revealed different types of bit wear in their premolars that are consistent with horse riding (*10, 17*). Second, equine lipid residue was identified in pottery at the Botai site, indicating animal husbandry and use of secondary products (*10*). Botai represents the earliest unambiguous evidence for domestic horse herding and riding (*17*), and, therefore, studying this population is essential for understanding the population dynamics surrounding horse domestication and determining the demographic impact of Botai in other prehistoric groups in which the horse was also a central cultural element. A more detailed description of the Botai site and discussion of the origins of the Botai culture can be found in reference (*15*).

Samples were taken from 3 different individuals for DNA extraction and analysis. 2 are genetically male, and 1 is a genetically female individual. The fact that all 3 individuals are genetically very similar increases the probability that these individuals accurately reflect Botai population rather than exogenous individuals present at the site through mechanisms like marriage. 2 of the samples were taken from crania curated in Petropavlovsk Museum, denoted as “Botai Excavation 14, 1983” and “Botai excavation 15.” Botai 14 has a calibrated radiocarbon date range from 3108–3517 cal BCE (2σ, UBA-32662) and Botai 15 from 3026–3343 cal BCE (2σ, UBA-32663). Unfortunately, there are no detailed osteological reports regarding these individuals. Botai 14 represents one of the male individuals from the multiple burials alongside many horses discovered in 1983. Botai 15 is an isolated find of a cranium.

The third individual to be sampled was recovered from excavations at Botai in 2016 with several of the authors of this paper present. Osteological and archaeological observations regarding this inhumation are presented below.

S1.1.1 Osteological analysis:

1. Inventory: Most of the skeleton was present for analysis. Notable elements that were missing included the right tibia and fibula as well as most of the left hand bones. The majority of the vertebrae and ribs were present, though fragmented, and some were displaced, notably the axis and atlas.

2. Preservation: The general bone preservation was poor varying between (Grade 4 and Grade 5) (*74*), most likely related to the shallow burial position and to some animal and root disturbance. Overall, the bone surface preservation was not good enough to identify some of the more subtle types of pathological lesions that may have been present (e.g., periosteal new bone formation).

3. Sex: The pelvis had a broad sub-pubic angle (*75*), the presence of a ventral arc (*76*), a sub-pubic concavity (*76*), a medial ischial-pubic ridge (*76*), and a preauricular sulcus (*77*). These features are suggestive of a female individual. However, the angle of mandible (*75*), mandibular ramus (*78*), and mental protuberance (*77*) were more indicative of a male; although the nuchal area (*77*) at the back of the skull was more female in nature. Overall, the morphological characteristics indicated that this was likely to be a female individual, and the genetics confirmed this sex determination.

4. Age: This individual was likely to be older than 45 years of age at time of death, based upon the morphological features of the pubic symphysis (*79*) and auricular surface (*80*). Analysis of dental wear (*81*) indicated that this individual was likely to be middle aged, indicating a slightly younger age of at least 35 years plus.

5. Stature: The female was estimated to be approximately 1.597 ± 0.042m based upon measurements extrapolated from the right radius (*82*) (the only long bone that had not suffered post-mortem fracture in the ground). The individual was relatively slight.

6. Pathologies: Spicules of very discrete new bone formation were evident in left and right maxillary sinuses and are likely to be indicative of sinusitis. The left maxillary first molar had been chipped during life and developed calculus, mineralized dental plaque, at the fracture surface.

S1.1.2 Archaeological context

1. The burial was a relatively shallow one, next to a house. The foot end was more deeply buried that the head end. The burial position was not one associated with any particular known burial rite and might be considered to be slightly haphazard, given that the leg positions were not the same in flexion and the right hand was hyperflexed back on itself.

2. A projectile point was recovered from approximately adjacent to the T6 vertebra. This point is of a form consistent with the Eneolithic and made out of a stone material commonly seen worked at Botai. The point was immediately adjacent to the skeleton but not embedded in bone. This point can be interpreted in three ways: (a) this is a victim of violence and the point is associated with their death but was embedded in soft tissue, (b) the point was a grave good, though there are no others, and it is in an abnormal location for that purpose, or (c) it is a Botai point that has only become randomly associated within the deposit.

3. Given the relatively high position in the ground, there was some disturbance of the burial by roots and animal burrows. The displacement of bones was most likely the result of burrowing.

4. Most animal bones in the immediate vicinity were horse bones, but there was also a femur of a European beaver (*Castor fiber* L.).

5. The only material culture associated was the projectile point of Botai type and the skeleton has been radiocarbon dated to a calibrated range of 3368–3631 cal BCE (2σ, UBA-32666), which puts it at the earlier end of the Botai culture range.

S1.2 Skeletal materials from Sholpan and Gregorievka

Samples from two Early Bronze Age (EBA) (~2200 BCE) skeletons from the vicinity of modern-day Pavlodar, in the River Irtysh region, were also taken. The Botai culture ends at the start of the 3rd millennium BCE. The following 800 years are then relatively poorly understood in this region, with a severe paucity of well-characterized and well-dated sites. However, there are many EBA sites that have been discovered in the last 10 years in the Pavlodar region, including many along the River Irtysh (*83*) Sholpan 4 and Gregorievka 2 are both EBA funerary sites with stone-lined inhumations in pit-graves under Kurgans (*84*). The Sholpan 4 skeleton has been radiocarbon dated to a calibrated range of 2468–2619 cal BCE (2σ, UBA-32664) and the Gregorievka 2 individual to 2037–2285 cal BCE (2σ, UBA-32665). The burial form is similar to the Yamnaya of the Pontic steppe, so it could represent migration of Yamnaya people into North Eastern Kazakhstan, replacing earlier Eneolithic populations (*27*). An alternative hypothesis would be that the EBA formed out of the Eneolithic populations of Northern Kazakhstan but adopted new burial rite forms, potentially through the spread of ideas rather than people.

S1.3 Okunevo

The Bronze Age Okunevo archaeological culture (~2500–2000 BCE) of South Siberia is characterized by complex burial traditions and art. Okunevo sites were found at the Minusinsk Basin, an area which includes both steppe and taiga environments and is surrounded by mountains. While some authors have suggested that the Okunevo may have descended from more northern tribes that replaced Afanasievo cultures in this region (*85*), others believe the Okunevo culture was the result of contact between local Neolithic hunter-gatherers with western pastoralists (*86*). A more extensive description of Okunevo archaeological sites can be found in reference (*15*).

For the present genetic study we choose 18 samples from 7 kurgans that represent both the Uybat and Chernovaya periods of the Okunevo culture. According to the archaeological data the oldest are 5 samples from Uybat V, kurgan 1, and Uybat III, kurgan 1 (*86*). Radiocarbon data on two of them (RISE 675 и RISE 677: 2600–2400 BCE) support their early dating. Other samples belong to the Chernovaya period: Okunev Ulus, Verkhniy Askiz, kurgans 1 and 2, Uybat V, kurgan 4, and Syda V, kurgan 3 (*86*, *87*). 8 radiocarbon dates of these samples are within 2300–1900 BCE. The only deviant dating of 2600–2400 BCE was obtained for samples of Syda V, kurgan 3.

S1.4 Baikal Hunter-Gatherers

For the current study, we have analyzed tooth samples from Lokomotiv, Shamanka, Ust'-Ida, and Kurma, ranging from the Early Neolithic (~5200 BCE) to the Bronze Age (~1800 BCE). In (*88*) the authors have put forward the following chronology for the prehistory of the Baikal region: Early Neolithic (5503±14 – 5027±33 BCE), Middle Neolithic (5027±33 – 3571±88 BCE), Late Neolithic (3571±88 – 2597±76 BCE), and Bronze Age (1726±34 – 1726±34 BCE). The archeological record of the region is marked by the absence of cemeteries during an interval of approximately 1,500 years, with

the suggestion of genetic discontinuity at the level of uniparental markers (*89*). In reference (*15*), a more complete description of the material culture of these sites across time is provided.

S1.4.1 Lokomotiv

The Lokomotiv cemetery (LOK) was initially discovered in 1897 during the construction of the Trans-Siberian Railway (*90*). The total area of LOK is estimated to be approximately 5,000 sq. m (*91*). The site is situated on a promontory at the junction of the Irkut and Angara rivers, approximately 70 km downstream of Lake Baikal, in a downtown park in Irkutsk (52°17.13.N, 104°14.57.E). Since its original discovery, LOK has been excavated on several occasions, mostly in conjunction with various construction projects carried out in and around the park. Between the 1920s and 1950s, 26 graves were excavated (*92*), but more systematic large-scale excavations were undertaken at LOK only during the 1980s and 1990s, uncovering 59 graves with a total of ~100 individuals (*18*, *19*, *91*). Some of these graves were excavated in the section of the cemetery referred to as Lokomotiv-Raisovet (LOR). The cemetery represents the Early Neolithic Kitoi mortuary tradition.

S1.4.2 Shamanka II

The Shamanka II cemetery (SHA) is located on the coast of Lake Baikal at its southwest end (51°41.54.N, 103°42.11.E). The cemetery is situated on a narrow peninsula that juts out into the lake in the E-W direction, between the small towns Sliudianka and Kultuk. The site was first discovered in 1962 when 3 graves were found to be eroding away from the cliff of the peninsula. No further fieldwork was done until the 1990s when 7 more graves were rescued from the collapsing cliff. During the 2000s, the cemetery has been excavated by BAP. Including the materials obtained in the 1990s, Shamanka II has produced 97 EN graves of the Kitoi mortuary tradition with about 155 individuals, 12 EBA graves of the Glazkovo mortuary tradition with 10 individuals, and 1 Late Bronze Age grave with 1 individual (*88*).

S1.4.3 Ust'-Ida I

The Ust'-Ida I cemetery (UID) is located on the bank of the Angara River at the mouth of its right tributary, the Ida, ~180 km north of Lake Baikal (53°11.20.N, 103°22.05.E). In the 1950s A. P. Okladnikov recorded 1 grave, and several more were spotted by amateur archaeologists in the mid-1980s (*93*). From 1987 to 1995, the cemetery was subjected to systematic archaeological excavations directed by V. I. Bazaliiskii (Irkutsk State University). This fieldwork produced 1 EN Kitoi grave, 31 LN Isakovo graves, and 19 EBA Glazkovo graves with 1, 47, and 20 individuals, respectively.

S1.4.4 Kurma XI

The cemetery of Kurma XI (KUR), comprised of 26 excavated graves, is located on the northwest coast of the Little Sea area of Lake Baikal, ~12 km northeast of the mouth of the Sarma River XIV cemetery (53°10.45.N, 106°57.46.E). One grave was excavated in 1994 by Irkutsk State Technical University, and the remaining 25 were excavated in 2002 and 2003 by BAP (*94*). Based on the typological criteria, 6 graves, all with poorly preserved skeletal remains, were classified as Late Mesolithic / EN and the remaining 20 as EBA Glazkovo mortuary tradition.

S1.4.5 Chronology

All human skeletal remains examined by BAP are also radiocarbon dated (*88*). The most recent round of this chronological research has included correction of the conventional ^{14}C dates for the freshwater reservoir effect and Bayesian modeling of various sets of dates (*88, 95–97*). Consequently, all individuals included in this study also have associated conventional, corrected, and calibrated ^{14}C dates presented in the Table S3 together with relevant archaeological and biological information. In most cases, radiocarbon dating confirmed the typochronological assessments.

S1.5 Anatolian materials

Kristian Kristiansen, Sachihiro Omura, Süleyman Yücel Senyurt, Fulya Eylem Yediay, Gojko Barjamovic

In this section we provide a compact overview of the skeletal material sampled for sequencing in the present work. For a more comprehensive summary of the main cultural phases of the Caucasus and Anatolia regions from 4000–1500 BCE, see reference (*48*).

S1.5.1 Kaman-Kalehöyük excavations (Kaman, Kırşehir, Turkey)

*Director: Dr. Sachihiro Omura, Japanese Institute of Anatolian Archaeology, Çağırkan, Kaman, Kırşehir, Turkey

The archaeological site of Kaman-Kalehöyük is situated in the Kızılırmak river basin in Central Anatolia. The main mound measures 280 m in diameter and is 16 m high.

The stratigraphy of the site can be divided into four major sections and several substrata:

1) Stratum I Ottoman/Islamic and Byzantine periods (1400–1600 CE)
   - Stratum Ia 1–3: Ottoman Period
   - Stratum Ib 4–5: Byzantine Period
2) Stratum II Iron Age and Hellenistic periods (1200–30 BCE)
   - Stratum IIa 1–2: Hellenistic Period
   - Stratum IIa 3–5: Late Iron Age
   - Stratum IIa 6–11: Middle Iron Age
   - Stratum IIc 2–3: Middle Iron Age
   - Stratum IId 1–3: Early Iron Age
3) Stratum III Middle and Late Bronze Age (2000–1200 BCE)
   - Stratum IIIa: Late Bronze Age (“Hittite Empire period”) (~1500–1200 BCE)
   - Stratum IIIb: Middle to Late Bronze Age (“Old Hittite period”) (~1750–1500 BCE)
   - Stratum IIIc: Middle Bronze Age (“Assyrian Colony period”) (~2000–1750 BCE)
4) Stratum IV Early Bronze Age (2300–2000 BCE)
   - Stratum IVa 1–4: Intermediate Period
   - Stratum IVb 5–6: Early Bronze Age(~2000–2300 BCE)

Context Stratum Ia (Ottoman Samples)

MA2195 (FEY1): HS 12-01, 12 07 24, South, Sector L Female, 35–45, Ottoman Ia

MA2196 (FEY2): HS 08-07, 08 07 17, North, Sector XXXV, Grid XLIX-48 (*99*), Provisional Layer 3 Juvenile, 7–8, Ottoman Ia

Context Stratum IIa1–2 (Hellenistic Period Samples)

The Iron Age levels at Kaman-Kalehöyük—including the Hellenistic period—can be divided into 4 architectural substrata from IIa (youngest) to IId (oldest). Substratum IIa can be divided into 5

chronological units based on ceramics. From youngest to oldest these are IIa1–2, IIa3–5, IIab–IIc1, IIc2–3, and IId1–3. In unit IIa1–2 (Hellenistic Period) both human and animal skeletons were found in pits. These fall into 3 different burial types: some containing only animal skeletons, others containing only human skeletons, and some with mixed human and animal skeletons.

Pit P1156 in the North Sector XV: a human skeleton was buried in a flexed position. Human and animal bones were apparently deposited together deliberately. Such burial features appear only in stratum IIa1–2 and may be correlated with a population change as well as possibly linked to incoming Galatians like at Gordion.

One of the pits **P1056** in sector XV also belongs to the Hellenistic Period.

MA2197 (FEY3): P1056 94 07 11 North, Sector XV, Grid XXXVI-52 (*5*) Provisional Layer 10, Hellenistic period. A skeleton of a juvenile aged 5–6 years came from P1056 was found alongside a small pig and four half-complete ceramic vessels.

MA2198 (FEY4): P1156 94 09 08, North, Sector XXVII, Grid XLVI-52 (*67*) B+C (Female C), Hellenistic Period, stratum IIa1–2

Context Stratum IIIb ("Old Hittite Period" Samples)

Based on findings, such as pottery and seals, stratum IIIb can be dated to the late part of the 2nd millennium BCE contemporary with the emergence of the Hittite state (1990 excavation reports).

MA2200-01 (FEY6): HS 89-01, 89 08 17, Sector III, Grid XLI-54 (C), Provisional Layer 48 – IIIb – Old Hittite Period. A partial skeleton was found in the west of section C together with an adult skeleton. Only the upper part of the first skeleton (skull, arms) was preserved (Kaman-Kalehöyük Field Notes 1994).

MA2203-04 (FEY8): 535 950810, North, Sector VI, Grid XXXIV-54 (M), Provisional Layer 61, Old Hittite Period. Skeleton HS95-35 belonging to a juvenile was found after removing room R141 on top. This layer is next to room 161, which is contemporary with stratum IIIb.

Context Stratum IIIc ("Assyrian Colony Period" Samples)

The Middle Bronze Age at Kaman-Kalehöyük represented by stratum IIIc yields material remains (seals and ceramics) contemporary with the international trade system managed by expatriate Assyrian merchants evidenced at the nearby site of Kültepe/Kanesh. It is therefore also referred to as belonging to the "Assyrian Colony Period" (*98*). The stratum has revealed three burned architectural units, and it has been suggested that the seemingly site-wide conflagration might be connected to a destruction event linked with the emergence of the Old Hittite state (*99*). The first burned architectural unit includes Rooms 148, 150, 298, 305, and 306. The second includes Room 153 and 208. The two units were excavated between 1994 and 2003. The third unit includes Room 367 and 370 and was excavated in 2004. Omura (*100*) suggests that the rooms could belong to a public building, and that it might even be a small trade center based on the types of artifacts recovered. Omura (*100*) has concluded that the evidence from the first complex indicates a battle between 2 groups took place at the site. It is possible that a group died inside the buildings, mostly perishing in the fire, while another group died in the courtyard.

MA2205 (FEY9): HS 11-1, 110705, North, Sector VIII, Grid XXX-55 (WW), Provisional Layer 75, Assyrian IIIc. Skeleton HS 11-01 was found in Sector (opening) VIII under a floor between Pit 1913 and Pit 3117 near pit 3117. It is thought to be a child based on its small size.

MA2206 (FEY10): 940826 **S1 (Skeleton1),** W4-W7 North, Sector I, Grid XLV-54 (GG) Provisional Layer 27, Assyrian IIIc. Room 153 belongs to one of the burnt architectural complexes that were excavated from Sectors 0, I, XXI, and XXII, and it is associated with the other burnt rooms dating to the Assyrian Colony period. Human skeletons were found between the exit of Room 153 and Wall 6 (Kaman-Kalehöyük Field Notes 1994).

MA2208-09 (FEY12): 940826, **S2 (skeleton 2)**, North, Sector I, Grid XLV-54 (GG), Assyrian IIIc. The sample comes from the same location as MA2206 above. There were 2 additional skeletons (S3 and S4) found here for a total of 4 individuals. They are thought to represent an opposing group fighting the individuals in Room 153. The skeletons fell on top of one other. They were not damaged by the fire.

S1.5.2 Ovaören excavations (Nevşehir, Turkey)

*Prof. Dr. Süleyman Yücel Şenyurt, Gazi University, Faculty of Arts, Department of Archaeology, Ankara/Turkey, Email: senyurt63@gmail.com

The multi-period archaeological site of Ovaören site is located in the Nevşehir Province, 20 km south of the Kızılırmak River. The site measures ~500 by ~350 m and consists of three areas main: Yassıhöyük (mound), Topakhöyük (mound), and its large terrace settlement (Fig. S2).

The main mound of Yassıhöyük was enclosed by a city wall 1250 m long during the Late Bronze Age (~1650–1150 BCE) and Middle Iron Age (~950–550 BCE). The Middle Iron Age layers represent a center in the region known as Tabal and belong to the Neo-Hittite cultural sphere (*101*). Later settlement on the mound dates to the Persian, Hellenistic, and Roman periods, but remains of these periods are mostly scanty.

Excavations conducted in 2013 on the terrace settlement beneath Topakhöyük revealed a number of skeletons in trench GT-137 from an Early Bronze Age context. The trench held 5 m of cultural deposit divided into 6 layers. Although no architectural evidence dated to the Middle Bronze Age was detected in the topmost layer (I), some trace of occupation was indicated by thrash pits that had been sunk into the Early Bronze strata from above. Two stone cist graves (M3 and M4) were found below 30 cm of cultural fill of layer I. Both lacked a cover slab, were empty, and probably robbed (*102*).

Layer II of GT-137 is represented by architectural remains as well as a mixture of Middle Bronze Age and Early Bronze Age pottery.

Layer III of GT-137 is characterized by large ash pits and scattered stones, especially at the eastern end of the trench, probably constituting a dump. An interesting feature in layer III was a planned cesspit 2 m wide by 2.5 m deep with an inner face created by a single line of stones. Finds, such as a tankard, *depas amphikypellon*, and sherds of wheel-made plates as well as Syrian Bottles date the stratum to the Early Bronze Age III

Layer IV of GT-137 likewise dates to Early Bronze Age IIIa based on architectural remains and finds, such as a bronze toggle pin, wheel-made plates, Syrian Bottles, and *depas amphikypellon*.

Layer V of GT-137 was the richest in terms of architectural finds and dates to the Early Bronze Age II. In this layer, 2 different structures and a well were uncovered. The well was filled with stones, pottery, and human skeletons (Figs. S2 and S3). In total, skeletons belonging to 22 individuals, including adults, young adults, and children, must belong to the disturbed Early Bronze Age II graves adjacent to the well (*103*). Pottery and stones found below the skeletons demonstrate that the water well was consciously filled and closed. The fill consists of dumped stones, sherds and skeletons, and the closing stones demonstrate that the water well was consciously filled and cancelled.

Samples from Ovaören-Topakhöyük:

MA2210: G-137, the well of layer V, individual no. 12.

MA2212: G-137, the well of layer V, individual no. 2.

MA2213: G-137, the well of layer V, individual no. 10.

S1.6 Turkmenistan samples

S1.6.1 Namazga samples

Vyacheslav Moiseyev, Andrey Gromov
Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), RAS.

Whereas most of current Turkmenistan was occupied by deserts during the Holocene, favorable climatic conditions and a good water supply in its southern part meant that agriculture appeared in the area ~5000 BCE. Most Eneolithic sites of Southern Turkmenistan are concentrated in the river valleys north of the Kopet Dag Mountains. The abundant natural flora and fauna in this area included wild fruit trees, wine, barley, sheep, and goats, which formed the basis for introducing agriculture and animal husbandry.

It is generally agreed that the Eneolithic of Southern Turkmenistan resulted from developments in the Neolithic Jeitun culture (*104*). Most sites of Southern Turkmenistan are multilayer settlements occupied from the Neolithic to the Bronze Age and later. The archaeological periodization of the Southern Turkmenistan Eneolithic is based on correlation of pottery types with cultural layers. In contrast to adjacent Neolithic cultures, Turkmenistan Eneolithic and later Bronze Age pottery were decorated with painted ornaments. The etalon periodization scheme was suggested by B. A. Kuftin and is based on a study of ceramic types from Namazga Depe and Anau settlements. This includes 4 Eneolithic pottery complexes of Anau 1a, Namazga I–III, and 3 Bronze Age complexes of Namazga IV–VI (*105, 106*). This scheme with several amendments is still in wide use.

The data on early agricultural cultures of Eastern Europe and the Caucuses suggest close interactions between early farmers and ancient pastoralists of the Eurasian steppe zone (*107*). In the case of Southern Turkmenistan, these would be Yamnaya, and later, Andonovo groups. The first evidence of influence of Yamnaya-Catacomb cultures adjacent to Turkmenistan territories was reported in the 1960s for the Zamanbaba burial site located in the Zarafshan area of modern Uzbekistan (*108*). This finding was proved by later excavations in the Zarafshan. At present, it is generally agreed that local Neolithic Kelteminar population of the Zarafshan area in the Eneolithic and later times maintained contact with both steppe pastoralists and early farmers of Southern Turkmenistan. Among the main features suggesting influence by Yamnaya (and possibly also Afanasievo) culture on local cultural traditions are such characteristics as single, crouched burials in simple pits graves or graves with a side grave chamber as well as pottery types characteristic to the steppe-zone cultures. Obvious Yamnaya influence in the area was further revealed by a study of the Zhukovo sacral complex 16 km from the city of Samarkand. It has been suggested that one of the main reasons behind the apparent expansion of Yamnaya into the Zarafshan was an abundance of metal resources in the area (*109*).

S1.6.2 Kara-Depe

The Eneolithic and Bronze Age settlement of Kara-Depe spanning the end of 5th to the beginning of the 3rd millennia BCE is located 4 km north of the village of Artyk, Akhalsk velayat, Turkmenistan (37.56/59.34). The site was first discovered by A. A. Maruschenko in 1930. It was excavated by B. A. Kuftin in 1952 and V. M. Masson in 1955–1957, 1960, and 1962–1963. The remains of the settlement formed a 15 ha mound, 11.5 m high. The depth of the cultural layers is estimated at 12.5 m.

The Early Eneolithic layers (Namazga I) of the late 5 millennium BCE are represented only by ceramics. For later layers of the same Namazga I period (early 4th millennium) remains of one-room square houses built of raw bricks with painted floors were reported. In the Middle Eneolithic (middle to end of the 4th millennium, Namazga II period), houses still had a single room, but their structure had become more complex. The room was usually divided into a number of sections and had a fenced yard.

Graves were in many cases lined with raw bricks. The deceased were usually buried lying on their side with bent legs. Numerous personal decorations were found, made of silver, gold, turquoise, lapis lazuli, and other kinds of ornamental stones.

The building structure grew more complicated in Late Eneolithic times (Namazga III). The settlement now consisted of one- or two-room houses with additional inner sections and additional compartments forming building blocks. Often these blocks were divided by narrow streets. Some of the most characteristic artifacts of the time are terracotta male and female figurines with complex relief details. While most of the pottery is characterized by monochrome black geometrical ornaments and animal representations of local origin, the presence of imported ceramics from the Southern-Eastern Caspian was also reported (*108*).

The two samples used for genetic analysis come from burials 42 and 43, matching layers of the Kara 2 (Namazga III period). The grave pit was located lower than the floor of the buildings of the Kara 2 layer, and cut through a Kara 3 cultural layer. The burial place is the largest on Kara-Depa and consisted of 35 inhumations. Graves were lined by raw bricks. Most of the skeletons lay on their right side with bent legs. Only a few pottery fragments were found in the graves.

See reference (*48*) for an in-depth contextualization of the Namazga and surrounding archaeological cultures.

S2: Ancient data analyses

Peter de Barros Damgaard*, Rui Martiniano*, Jack Kamm*, José-Victor Moreno-Mayar*, Arjun Biddanda, John Novembre, Rasmus Nielsen, Martin Sikora, Richard Durbin**, Eske Willerslev**

* contributed equally

** corresponding authors

S2.1 Data generation

74 ancient genomes were generated using state-of-the-art processing of ancient skeletal material: targeting petrous bones or tooth cementum, extracting and building NGS libraries according to approaches described elsewhere (*1, 110*). However, we coupled these advances to a novel NGS approach by sequencing ancient DNA libraries on the Illumina X10 platform, hereby reducing the sequencing cost considerably. The geographical location of the ancient samples sequenced in the present work is represented on Fig. 1 in the main paper, where we also define the boundaries of western, central, and eastern steppe regions (terrestrial ecoregions shapefile data downloaded from the Nature Conservancy, http://maps.tnc.org/). We note that these are present-day geographical limits and may not correspond exactly to the distribution of steppe regions in prehistory.

Briefly, teeth or petrous bones were drilled for either well-preserved cementum or compact otic capsule bone, in the dedicated clean-laboratories (*111*) of the Centre for GeoGenetics at the University of Copenhagen. The drilled samples were then decontaminated efficiently removing bacterial and fungal DNA using a 30 minute pre-digestion (*110*) slightly modified to consist of 4.9 mL EDTA and 100 uL Proteinase-K. The DNA was then extracted from the solution using a modified Qiagen PB Buffer binding buffer developed in (*1*) for binding ultra-short DNA sequences and eluted in 82 uL commercial EB Buffer. Then, 3–4 standard Illumina next-generation sequencing libraries were built per extract using 20 uL extracts per library, according to a modified NEB Next protocol (*112*). These were amplified using a pool of 4 indexes per library, thus providing the required base complexity for the sequencing of single libraries per lane on the Illumina platform, hereby circumventing “index bleeding”

characteristic of the X platform (*113*). For all libraries, the Kapa U+ enzyme was used for amplification due to its low GC-bias (*114*), and all libraries were amplified for 14–18 cycles. Libraries were sequenced in single read mode at the Danish National Sequencing Center using an Illumina HiSeq 2500 to 80 bp, and in paired end mode, 151 cycles (302 cycles total) at the Wellcome Trust Sanger Institute, Hinxton, UK.

In addition to the 74 ancient genomes presented in this study, we also sequenced 41 high-coverage genomes (30X) on the Illumina X10 platform in South Korea (Fig. S5). We merged this novel data with high coverage genomes from previous studies (*1, 115*). For exhaustive description of genotyping parameters see **Section S2.6**. All saliva samples used for generating high coverage genomes were collected by a close collaborator of the Eske Willerslev research group complying with legal requirements. All donors provided informed written consent stored in Copenhagen. Permission for undertaking the study in the country of the corresponding author in Denmark was obtained according to the Danish National Committee who deemed the study non-notifiable according to the Committee Law paragraph 14. The samples were all anonymized and remain identifiable only by the first author.

In addition, we genotyped 140 individuals from 5 populations in Pakistan (Gujar, Kohistani, Tarkalani, Uthmankhel, and Yusufzai), using the Infinium OmniExpressExome-8 v.1.3 BeadChip array platform. All samples were collected by a member of the Eske Willerslev research group for demographic analyses in the districts of Swat and Dir. All donors provided informed written consent, and permission for undertaking the study in the country of the corresponding author was obtained according to the Danish National Committee who deemed the study non-notifiable according to the Committee Law paragraph 14. We merged this novel data with genotype data from present-day Indian populations (*43*) and with the merged dataset from (*3*), which is enriched in individuals with Eurasian ancestry from various time periods ranging between the Mesolithic to the present. The merged dataset consisted of 236811 SNP sites for 1805 individuals from 165 populations.

S2.2 Raw read processing and mapping

We converted CRAM files containing paired-end sequencing data to interleaved fastq using samtools (*116*), removing sequences that fail platform and vendor quality checks. Adapter sequences were trimmed using AdapterRemoval2 (*117*), collapsing overlapping read pairs, trimming Ns and low quality bases (quality threshold 2) as well as selecting reads with minimum length of 30. Single read data was also trimmed using AdapterRemoval2 with the same parameters, except for read collapsing and interleaved input options. Next, we aligned truncated reads to the reference genome hs.build 37.1 using bwa aln (*118*) -l1024 and bwa samse, and used samtools (*116*) to keep mapped reads with mapping quality equal or above 30. Read duplicates were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/), and we added read groups to reads with AddOrReplaceRG. We merged bam files belonging to the same sample, which we then processed with the Genome Analysis Toolkit (GATK) Target Creator (*119*), providing known indels from the 1000 Genomes followed by Indel Realignment. Finally, we used samtools calmd to generate the MD tag with extended BAQ calculation. Genomic coverage was calculated using qualimap with default parameters (*120*). We present basic sequencing statistics and post-mortem DNA damage in Table S1.

S2.3 Contamination estimates

We estimated contamination using two approaches: first, using contamMix (*121*), an approach that compares the mapping affinities of each mitochondrial read to the consensus sequence of the individual

with the mapping affinities to worldwide dataset of putative contaminants assembled in (*122*). This approach can be used successfully on all individuals with a mitogenomic coverage > 10X. Secondly, we estimated contamination using a method developed for males in (*123*) implemented in ANGSD, taking advantage of variation at the X-chromosome to assess contamination. We show estimated contamination values in Table S1.

S2.4 Sex determination

We used the Rgamma statistic, i.e., the number of sequences mapping to the Y chromosome divided by the total of number of sequences mapping to sex chromosomes (*124*) to determine the sex of these ancient individuals (Table S1).

S2.5 Relatedness

Including relatives in population frequency-based statistics could lead to incorrect assessments. Secondly, related individuals may be informative for interpretation on social organization. For these reasons, we estimated relatedness between all pairs of individuals using a two-step approach. We first calculate all the outgroup-f3 statistics of the form f3(Individual X, Individual Y; Mbuti) in order to identify and flag pairs of individuals with inflated levels of shared ancestry (Fig. S6). To follow up on this method, we estimated biological relatedness between pairs of individuals using LCMLKIN (*125*) (https://github.com/COMBINE-lab/maximum-likelihood-relatedness-estimation). An advantage of LCMLKIN is to use genotype likelihoods instead of genotypes and therefore not assuming that genotypes are ascertained without error. This is of particular importance in ancient DNA studies, where low coverage data is abundant.

First, we selected 300,000 SNPs at random from the Human Origins dataset (*42*). Next, we called genotype likelihoods at these SNP positions using 'SNPbam2vcf.py' provided with LCMLKIN. Finally, we estimated biological relatedness between pairs of individuals using LCMLKIN. Individuals with high relatedness are shown on Table S5.

Having verified that a large number of Okunevo_EMBA pairs present high levels of relatedness and given that we sampled individuals from 4 distinct burial sites (Syda 5, Uybat, Okunev Olus, and Verkhni Askiz), we wanted to investigate whether these values represented mobility across different communities or instead were the result of temporal and geographic proximity within communities. We plotted pairwise coefficients of relatedness according to geography (Fig. S6) and verify that the highest values were obtained between individuals belonging to the same burial site, in particular those of Verkhni Askiz and Okunev Olus, and we do not see exceptional values of affinity between individuals from different sites. Specifically, the highest values obtained were for individuals belonging to the Verkhni Askiz population with 2 pairs of individuals showing pi_HAT of 0.41 (RISE516-RISE672) and 0.48 (RISE515-RISE673) which may imply these are first-degree relatives. Additionally, possible second-degree relatedness, with values around 0.2 were also identified in Verkhni Askiz, but also between 1 pair of Okunev Ulus individuals. The likely explanation for the high relatedness observed between Verkhni Askiz individuals is that they were retrieved from only 2 directly adjacent burials with a span of a mere 100 years. In contrast, the remaining burials span ~400 years.

Four pairs of individuals from the Baikal Lake region also presented high coefficients of relatedness, with each pair of individuals belonging to the same archaeological site: Shamanka_EBA (DA336 and DA338, pi_HAT=0.589; DA334 and DA335, pi_HAT=0.388), Lokomotiv_EN (DA340 and DA341, pi_HAT=0.290), and UstIda_EBA (DA353 and DA361, pi_HAT=0.240). Lastly, high

relatedness was also detected in two Namazga_CA samples (DA379 and DA380) which presented a pi_HAT=0.458.

S2.6 Genotyping

All genomes were genotyped individually using samtools (v1.3.1) mpileup –C50 and bcftools (v1.3.1) using the consensus caller (*116*). Calls from each genome were filtered for a minimum of 1/3 average depth and a maximum of 2 times average depth, except for the mitochondrial genome, which were filtered for a minimum 10 and maximum 10000 read depth. For males the X and Y chromosome were filtered using half the threshold as for the autosome. The variant calls were subsequently filtered if there were two variants called within 5 nt of each other, for phred posterior probability of 30 and strand bias, end distance bias of p<1e-4 and read position bias of 0. Additionally, we filtered heterozygote sites if allelic balance for the minor allele was less than 0.25. Per individual calls were merged across all samples using GATK-3.7 CombineVariants (*119*) to per chromosome files and filtered for deviations from Hardy-Weinberg Equilibrium with p>1e-4 (*126*).

S2.7 Principal Component Analysis

We carried out the PC Analyses on different subsets of populations using 236811 SNP sites previously filtered in *(3)*. These include:

- the full Eurasian panel described in (*3*), including the novel 74 ancient genomes (Fig. S13; Fig. 2A)
- a subset of the Eurasian panel described in (*3*), including the novel 74 ancient genomes and the South Asian populations from (*43*) (Fig. 2B) focusing on the major gradients defining South Asian ancestry
- a subset of the Eurasian panel described in (*3*) focusing on relevant modern populations from the Altai and Siberia and the ancient genomes (Fig. S23) defining the ANE-to-AEA genetic cline.

We used PLINK 1.9 (*127*) to perform Principal Component Analyses including the ancient samples in the calculation.

S2.8 Model-based clustering

We computed model-based clustering analyses on the Eurasian panel in order to explore shared ancestries between the past and present groups. For each K = 2 to K = 15 we computed 20 replicates and we show the admixture proportions for all ranges of K in Fig. S14. For each value of K, we estimated the 5-fold cross-validation error based on the maximum-likelihood solution across replicates (Fig. S15). We observe minimum cross-validation error estimates when assuming 6 and 10 ancestral populations. We show admixture proportions for K = 6 in the main text.

S2.9 D-statistics

We computed allele frequency-based *D*-statistics (with *AdmixTools*) to formally test hypotheses about the ancestry composition of different groups in the merged dataset. In brief, *D*-statistics of the form *D*(*H1*, *H2; H3*, *H4*) are expected to be consistent with 0 if *H1* and *H2* form a clade in the unrooted tree (((*H1*, *H2*), *H3*), *H4*). Significant deviations from this expectation may arise due to the proposed tree being wrong, gene flow between the lineages in the tree, or differential error rates

between *H1* and *H2*. In order to assess the statistical significance of the deviation, we estimated the standard error for each statistic using a weighted block jackknife approach over 5Mb blocks and computed *Z*-scores for each value of *D*. We consider *D*-statistics for which $|Z|>3.3$ (p-value < 0.001) to be significantly different from $D = 0$. Since different groups bear variable error rates mostly derived from post-mortem DNA modifications, we performed this analysis on the complete merged dataset, and a filtered version where we discarded transition polymorphisms.

S2.10 qpAdm modeling

S2.10.1 Methods

Following the results presented in previous sections and in the main text, we modeled the admixed ancestry of a set of modern and ancient populations using qpAdm (*2*), as implemented in *AdmixTools* latest version. This method models a "target" population as a mixture of *n* different "source" populations, which are differentially related to a set of *m* different "outgroups." Thus, $f4(Target, Outgroup_j; Outgroup_k, Outgroup_l)$ can be expressed as a weighted sum of all possible statistics of the form $f4(Target, Outgroup_j; Outgroup_k, Outgroup_l)$. Additionally, qpAdm provides a test for the proposed model via qpWave. This test is meant to assess whether the target and *n* source populations derive from at least *n* independent "migration streams" from the *m* outgroups. Therefore, for each of the proposed models, we first tested if the selected set of outgroups were informative about the different ancestries of a given set of source populations. We tested each model on both the full merged dataset and on a dataset filtered for transition polymorphisms.

S2.10.2 Assessing outgroup informativeness

For each of the qpWave models described in the main text, we used the following set of outgroup populations genomes:

Ust_Ishim
Anzick1
Kostenki14
Switzerland_HG
Natufian
Mal'ta (MA1)

Since qpAdm assumes that the source populations are differentially related to the outgroups, we first assessed whether this set of outgroups was informative about the different ancestries carried by the sources. We first computed all possible statistics of the form $f4(Target, Outgroup_j; Outgroup_k, Outgroup_l)$. If a pair of potential sources is equally related to the outgroups, we expect the f_4-statistics for this pair to be highly correlated; thus, suggesting that the outgroups are not informative about such sources (*42*). While we did not find any of the source pairs to yield near perfectly correlated statistics (Fig. S36), pairs such as (CHG, IranN) yielded correlation scores as high as 0.92 indicating that these ancestries might not be optimally identified using our set of outgroups with this approach. In addition, we note that the power will be lower when trying to differentiate between the following pairs:

(Namazga, IranN), cor ~ 0.929
(Namazga, CHG), cor ~ 0.948
(Namazga, Turkmenistan_IA), cor ~ 0.938
(Steppe_MLBA, Steppe_EMBA), cor ~ 0.91
(CHG, IranN), cor ~ 0.929

For the remaining sources, this test suggests that the set of seven outgroups allows us to confidently differentiate between the different proposed sources. For each model in the main text, we confirmed these results by assessing if the source populations in turn could be expressed as independent "migration streams" from the outgroup populations using qpWave (Table S16). For all models, we found statistically significant evidence (p-value < 0.05) for the source populations to be differentially related to the outgroup populations. When filtering out transition polymorphisms, we found non-significant qpWave p-values (Table S17), yet we interpret these results as a consequence of reduced statistical power due to the low number of remaining SNP positions.

S2.11 qpGraph shows no evidence of Botai-Yamnaya gene flow

To validate our finding of no Botai-Yamnaya admixture, we used qpGraph (Admixtools https://github.com/DReichLab/AdmixTools) to fit a simple admixture graph on Yamnaya, Botai, EHG, CHG, Xiongnu (representing East Asian ancestry), and Mbuti (outgroup), using transversion SNPs and a jackknife block size of .05 Morgans. This graph (Fig. S28) had no direct Botai-Yamnaya gene flow and fit all f4 statistics (|Z| <= 1.77), agreeing with other results that show no evidence of direct gene flow between Yamnaya and Botai.

S2.12 Chromopainter

We extracted from our call set 621,799 positions genotyped in the Human Origins dataset (*42*). We merged variants in our call set with the Human Origins genotype dataset using PLINK 1.9, and filtering for missingness per individual (--mind 0.51) and missingness per marker (--geno 0.05), resulting in a total of 1,250 individuals genotyped for 581,755 SNPs, including the newly sequenced ancient samples BOT2016 (Botai), Sholpan (Central Steppe EMBA), and Yamnaya Karagash, and the previously published Ust-Ishim (*128*). We then used SHAPEIT v2.r790 (*129*) in mode "check" to detect variant alignment errors in our data, which we excluded from the dataset, resulting in 540,070 SNPs. We subsequently phased these genotypes using SHAPEIT with default parameters, providing the 1000 Genomes Phase 3 haplotypes and recombination map as a reference (http://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3/). Next, we converted phase files with impute2chromopainter.pl and converted the 1000 Genomes recombination map with convertrecfile.pl into the format required by fineSTRUCTURE. Both of these scripts were downloaded from http://www.paintmychromosomes.com/.

We used fineSTRUCTURE v2 (*35*) (https://people.maths.bris.ac.uk/~madjl/finestructure/) to investigate patterns of haplotype sharing in our data. We examined the "chunkcounts" output file produced in our analysis above and estimated the mean haplotype sharing with present-day populations and each one of the 3 newly sequenced high-coverage ancient samples (Fig. S29).

Consistent with previous reports of mass migration of steppe pastoralists into Europe (*1, 2*), the Yamnaya sample shows a substantial contribution to present-day Europeans, in particular Karelians and Ukranians. Conversely, Botai shows higher affinity to Yeniseians, Native Americans, Eskimos, Tubalars, Selkups, and other Far Eastern Siberian populations. The affinity between Botai and Eastern and Northern European groups is non-negligible, however when interpreted together with results from other analyses presented in the manuscript, in which we report Botai's ancestral link to ANE, the observed sharing patterns are likely to derive from the MA1-related ancestry it shares with Yamnaya, rather than from a direct contribution. Furthermore, the intensity of haplotype affinity shared by Yamnaya and West Eurasians is greater than that of Botai to Native Americans, Siberians, or any other

population, which suggests that the first horse domesticators contributed less to the genetic pool of modern populations than the Yamnaya, who have used the horse as a vehicle to spread into West Eurasia. The Early Bronze Age Sholpan sample presented haplotype affinity patterns broadly similar with Botai, with greatest affinity to the Yeniseian and Native American populations, but it is characterized by lower affinity to Europeans. To compare sharing patterns between the 3 samples, we normalized mean haplotype sharing values with present-day populations and present these in a ternary plot (Fig. S30). At the macro population level, Yamnaya has greater sharing with West Eurasians, while both Botai and Sholpan share more haplotypes with Native Americans and Eastern Eurasian populations, but with the latter sample showing greater proportions of Siberian and East Asian ancestry.

To allow for a more detailed comparison at the population level, we plotted pairwise comparisons between Yamnaya and Botai (Fig. S31A) and between Botai and the Sholpan sample (Fig. S31B) and estimated their correlation. Sholpan's patterns of mean haplotype sharing are more correlated with Botai's ($r = 0.58$), and this value is greater than the correlation between Botai and Yamnaya ($r = 0.51$). This may imply that despite ANE ancestry being present at different levels in these samples, both Sholpan and Botai are more related to MA1 than Yamnaya is, and that Yamnaya contains CHG ancestry, which further differentiates it from the 2 samples. In this detailed comparison, Sholpan shows greater affinity with certain Far Eastern populations than Botai, in particular with the Eskimo, Koryaks, Chukchis, and Yakuts as well as with Altai populations and Mongolic-speaking peoples.

To examine geographic differences in haplotype sharing with present-day populations between Botai and Yamnaya, we estimated the total variation distance statistic (*130*) (Fig. S32). The size of the circles highlights the magnitude of differences, while the color represents total contribution. We observe that Botai and Yamnaya differ in the amount of sharing with East Asians, with Botai showing higher values, but that the overall sharing of Botai and East Asians is very reduced, indicating small proportions of East Asian related ancestry in Botai not present in Yamnaya, consistent with the cline of ancestry shown on Fig. 2. On the other hand, with Native American populations, we observe large magnitude differences between Yamnaya and Botai, but, in this case, Botai shares a substantial amount of haplotypes with these populations.

S2.13 SFS-based modeling

In this supplement we describe how we used the site frequency spectrum to infer the model in Fig. 4 of the main text.

We followed a strategy of fitting a succession of increasingly complex demographic models. In particular, we fit the following models: (a) a small model for the demographic history of Yamnaya ancestry, (b) a slightly larger model for 3 central Eurasian steppe populations and a Baikal population, and (c) a large, 10-leaf model based on combining the first two models.

Our demographic models consisted of samples from 10 populations: YamnayaKaragash_EBA, SidelkinoEHG_ML, Botai_CA, CentralSteppe_EMBA, Okunevo_EMBA, MA1, KK1, Shamanka_EN (Lake Baikal), Mbuti, and Han. For YamnayaKaragash_EBA, Botai_CA, and CentralSteppe_EMBA, we used a single sample, excluding the low-coverage samples with less than 9x coverage. KK1 also consisted of a single ancient sample. We used 2 samples each from the modern Mbuti and Han populations.

MA1, SidelkinoEHG_ML, Okunevo_EMBA, and Shamanka_EN each consisted of only low-coverage samples (less than 9x coverage). For each low-coverage sample, we chose a random allele at each SNP where there was at least 1 read with mapping quality ≥33. While SidelkinoEHG_ML and MA1 each consisted of a single sample, Okunevo_EMBA and Shamanka_EN contained many samples; to speed up the likelihood computation, we downsampled each SNP to have 4 random alleles from

these populations. To adjust for the fact that we did not ascertain SNPs within the low-coverage samples, we only considered SFS entries that were polymorphic within the high-coverage samples and adjusted the denominator of the SFS so that all entries represented *conditional* probabilities, conditioning on the high-coverage samples being polymorphic.

In the remainder of this supplemental section we will usually refer to these populations by shortened names, so that they fit more easily in the figures. These shortened names are “Yamnaya”, “Sidelkino”, “Botai”, “Sholpan”, “Okunevo”, and “ShamEN”. Sholpan is the site of the 9x-coverage CentralSteppe_EMBA sample; the other shortened names are self-explanatory.

We used the method momi (*28*) to compute expected SFS values under the multipopulation coalescent, which were then combined into a composite likelihood, where the observed SFS was modeled to be drawn from a multinomial distribution, while the total number of heterozygotes per individual were modeled as independent Poisson variables (we used heterozygotes per individual, rather than the total number of SNPs in the dataset, because it is easier to account for the effect of missing data). Demographic models were then inferred by performing gradient descent to maximize this composite likelihood. To estimate confidence intervals, we used the parametric bootstrap with 300 simulations. We also used the parametric bootstrap to estimate the bias and standard deviation of our estimates.

For all models, we assumed a generation time of 29 years, and a mutation rate of 1.66×10^{-8} per base per generation, based on 2 recent estimates of the mutation rate (*131*, *132*).

S2.13.1 A simple model for Yamnaya ancestry

We began by fitting a simple 4-population model relating KK1, Sidelkino, Botai, and Yamnaya, shown in Fig. S16. The model included the following population admixture and split events:

1. An admixture event, where Yamnaya is formed from a CHG population related to KK1 and an ANE population related to Sidelkino and Botai. We inferred 54% of the Yamnaya ancestry to come from CHG and the remaining 46% to come from ANE.

2. A split event, where the CHG component of Yamnaya splits from KK1. The model inferred this time at 27 kya (though we note the larger models in Sections S2.12.4 and S2.12.5 inferred a more recent split time).

3. A split event, where the ANE component of Yamnaya splits from Sidelkino. This was inferred at about about 11 kya.

4. A split event, where the ANE component of Yamnaya splits from Botai. We inferred this to occur 17 kya. Note that this is above the Sidelkino split time, so our model infers Yamnaya to be more closely related to the EHG Sidelkino, as expected.

5. An ancestral split event between the CHG and ANE ancestral populations. This was inferred to occur around 40 kya.

We found that specifying a separate population size along each branch led to an over-parametrized model, with identifiability issues and runaway behavior. We thus fit a model with 4 population sizes:

1. A population size along the Botai leaf branch.
2. A population size along the KK1 leaf branch.
3. An ancestral population size at 100 kya.
4. A shared “Eurasian” effective population size along all other internal branches.

We summarize the inferred parameters, along with bootstrap estimates of bias, standard deviation, and 95% confidence intervals, in Table S6. In Fig. S17, we plot the bootstrap distribution of the difference in split times between Yamnaya and Botai/Sidelkino and can reject the hypothesis that Yamnaya split from Botai after Sidelkino at 95% confidence level.

S2.13.2 No significant Botai-Yamnaya gene flow detected

We used 2 approaches to investigate whether we could detect additional gene flow from Botai to Yamnaya related to the spread of horse domestication. First, we added extra pulses between Botai and Yamnaya and checked whether the inferred pulse strength was significantly different from 0. Second, we checked whether the model without gene flow could adequately fit statistics of excess allele sharing between Yamnaya and Botai. In both approaches, we found no significant signal of gene flow between Botai and Yamnaya.

In the first approach, we tried adding additional pulses between Botai and Yamnaya and re-estimating the MLE (Fig. S18). When adding a Yamanaya->Botai pulse, we inferred no gene flow (pulse strength of 0%). Adding a Botai->Yamnaya pulse, our model inferred a small amount of gene flow (pulse strength of 4.8%), but this was not significantly different from 0 (p-value .18) under 300 parametric bootstraps simulated under the null model without admixture.

In the second approach, we used a modified version of Patterson's "ABBA-BABA" f_4 statistic (*133*) to test for significant excess sharing between Botai and Yamnaya. In particular, drawing a single random allele from each of 4 populations P_1, P_2, P_3, P_4, let BABA be the number of SNPs where $P_1 = P_3 \neq P_2 = P_4$, and similarly let ABBA be the number of SNPs where $P_1 = P_4 \neq P_2 = P_3$. Then $f_4 = \frac{BABA - ABBA}{N}$ is the difference in the BABA and ABBA counts, normalized by some appropriate constant N. If the populations are related by the unrooted topology $((P_1, P_2),(P_3, P_4))$, then $f_4 \gg 0$ indicates excess BABA-type incomplete lineage sorting, due either to admixture between P_1 and P_3, or between P_2 and P_4.

f_4 is simply a statistic of the SFS, and so we can check whether the f_4 statistics of the observed SFS match the f_4 statistics of the expected SFS. Note this is similar to the approach of qpGraph (*133*) for checking whether f_4 statistics of admixture graphs match the data. However, qpGraph assumes that mutations are old and occurred in the root population, and it requires SNPs to be ascertained within an outgroup; whereas here we consider the effects of all SNPs, including those from recent mutations.

To check for admixture between Botai and Yamnaya, we compared ABBA-BABA counts for quadruples (Yamnaya, Sidelkino; Botai, *X*), varying the value of *X*. A relative excess of BABA counts (compared to the model expectation) indicates excess allele sharing between Botai and Yamnaya that is not shared by Sidelkino. However, instead of using the usual f_4 statistic, which is based on the difference of BABA and ABBA counts, we used a modified version of it, which we denote by f_4^*, and define as

$f_4^* = \log(\text{BABA}) - \log(\text{ABBA}) = \log(\frac{BABA}{ABBA})$.

That is, instead of using the difference of BABA and ABBA counts, we use the difference of their logarithms. f_4^* is robust to certain biases that may affect $f_4 = \frac{BABA - ABBA}{N}$ through the normalization constant N (the total number of observed SNPs). In particular, missing data or reference bias may cause a decrease in observed singletons, especially in lower-coverage individuals, leading to a decrease in the total number of SNPs. By contrast, f_4^* only depends on BABA and ABBA counts, which require 2 copies of each allele and thus are not affected by singleton counts.

To compute the empirical f_4^*, we counted the number of BABA and ABBA SNPs in every subsample of 4 alleles and took the log-ratio. To compute residuals, we compared this with the log ratio of BABA and ABBA probabilities, dividing by the standard error of f_4^* under a block jackknife with 100 blocks. We denote this normalized residual by Z^*, so

$$Z^* = \frac{\log(BABA/ABBA) - \log(P(BABA)/P(ABBA))}{SD(\log(BABA/ABBA))}.$$

For the model in Fig. S16 (without Yamnaya-Botai admixture), we found that the residuals of f_4^*(Yamnaya, Sidelkino, Botai,X) were not significantly positive for $X \in$ {KK1, AncestralAllele}, as shown in Table S7.

In addition, in Sections 2.13.4 and 2.13.5 below, we consider larger models that includes 6 additional populations (Mbuti, MA1, Sholpan, Shamanka EN, and Han), shown in Figs. S19 and S20. Most notably, these models account for East Asian ancestry in Botai, which is not considered in the model in Fig. S16. We checked the residuals of f_4^*(Yamnaya, Sidelkino, Botai,X) for these larger models; none of these residuals were significantly positive (Tables S9 and S11), consistent with a model of no recent genetic admixture between Botai and Yamnaya.

S2.13.3 Modeling the central Eurasian steppe 5,000 years ago

We next examined 3 related populations from the central Eurasian steppe 4–5.5 kya (Botai, Sholpan, and Okunevo), as well as an Ancient East Asian (AEA) population from Lake Baikal 7 kya (Shamanka Early Neolithic). For this model, we also included modern Mbuti and Han samples as well as the ancient MA1 sample from Siberia 24 kya.

We modeled the 3 steppe populations as a mixture of ANE and East Asian ancestry but with Botai having more ANE ancestry than the Okunevo and Sholpan samples. We based this model on several exploratory models for subsets of these populations, as well as PCA and qpAdm results that showed these 3 steppe populations to be closely related and intermediate between ANE and East Asian samples.

More specifically, we modeled the 3 steppe populations as splitting off from a "ghost" ANE population at time $T_{\text{Steppe-GhostANE}}$, and receiving a pulse of East Asian ancestry at time $T_{\text{AEA->Steppe}}$ shortly thereafter. We modeled this East Asian pulse as coming off the ShamankaEN branch. Later, the Botai population is formed at time T_{Botai} from an additional admixture event between the Steppe and GhostANE, while the Okunevo and Sholpan populations split from each other at $T_{\text{Sholpan-Okunevo}}$.

Additional split times in the model are $T_{\text{Han-ShamankaEN}}$ for the split between Han and ShamankaEN, $T_{\text{MA1-GhostANE}}$ for the split between MA1 and GhostANE, $T_{\text{AEA-ANE}}$ for the split between East Asian and ANE populations, and $T_{\text{Mbuti-Eurasia}}$ for the split between Mbuti and Eurasian populations. For the population size parameters, we generally inferred separate population sizes at leafs with high-coverage samples, while sharing population size parameters at low-coverage leafs with internal branches. Specifically, the high-coverage samples in Mbuti, Botai, Sholpan, and Han have effective sizes N_{Mbuti}, N_{Botai}, N_{Sholpan}, N_{Han}, respectively, while ShamankaEN and the ShamankaEN-Han ancestor have size N_{Han}, Okunevo and the Botai-Okunevo-Sholpan ancestor have size N_{Steppe}, MA1 and GhostANE have size N_{ANE}, the AEA-ANE ancestor has size N_{Eurasia}, and the Mbuti-Eurasian ancestor has size $N_{\text{Ancestral}}$.

We show the inferred maximum-likelihood model in Fig. S19 and bootstrap confidence intervals in Table S8. Specifically, we inferred the steppe populations to have 51% East Asian ancestry and 49% ANE ancestry, with Botai having an additional pulse of 40% ANE ancestry (for a total of .49 + .51 · .4 ≈ 0.69 of Botai ancestry coming from ANE). We inferred the admixture and divergence events relating Botai, Sholpan, and Okunevo to occur ~10–13 kya and inferred the divergence of ShamankaEN from Han ~17.5 kya.

S2.13.4 Combining the Yamnaya and central steppe models

We next constructed a large, 10-leaf model that combined the Yamnaya-focused model of Fig. S16 with the central Eurasian steppe model of Fig. S19. We show this model in Fig. S20.

More specifically, we constructed this model by starting with the model in Fig. S16, then adding on the Sidelkino, KK1, and Yamnaya leafs. Yamnaya was modeled as a mixture of Sidelkino with a CHG population related to KK1. We found the likelihood surface for the time of this admixture to be very flat, so we did not estimate this parameter, simply fixing it to occur at the time of the Yamnaya sample.

Compared to the central steppe model in Fig. S19, the divergence time of the central steppe populations decreased slightly, as did the MA1 divergence time; however, the Han-ShamankaEN, ANE-AEA, and Mbuti-Eurasian divergence times remained essentially the same. The ANE and AEA admixture proportions within the central steppe populations also changed by about 5 to 10%. Compared to the Yamnaya focused model in Fig. S16, the KK1-Yamnaya divergence time decreased to about 20 kya, but the KK1-ANE divergence time remained about the same (at ~40 kya), and the Yamnaya admixture proportions also remained essentially the same.

As discussed in Section S2.13.2, we checked whether there was excess Yamnaya sharing with Botai not accounted for by Sidelkino by examining the ratio of ABBA-BABA counts. The f_4^*(Yamnaya, Sidelkino, Botai, X) statistics (as defined in Section S2.13.2) are listed in Table S10. None of these f_4^* statistics was significantly positive, consistent with a model of no recent genetic admixture between Botai and Yamnaya. However, Z^*(Yamnaya, Sidelkino, Botai, Okunevo) $\ll 0$, suggesting excess allele sharing between Yamnaya and Okunevo, which agrees with both qpAdm results suggesting a Yamnaya-like contribution to Okunevo, and the geographic proximity of Yamnaya-related Afanasievo settlements to subsequent Okunevo settlements.

S2.13.5 Adding a Yamnaya->Okunevo pulse

Based on the Z^*(Yamnaya, Sidelkino, Botai, Okunevo) $\ll 0$ statistic in Table S9 as well as parallel lines of evidence from qpAdm and archaeology, we added a pulse from the Yamnaya to Okunevo leafs, resulting in the model in Fig. S21. The model inferred a 16% contribution from Yamnaya to Okunevo. The MLE point estimate and 95% parametric bootstrap confidence intervals are summarized in Table S10. We also show the f_4^*(Yamnaya, Sidelkino, Botai, X) statistics in Table S11; none of these statistics was significantly different from 0 at the 95% level after a Bonferroni correction.

S2.13.6 Robustness of results to errors in medium-coverage ancient samples

A possible complication of fitting the SFS with an explicit coalescent model is that the SFS can be affected by damage, such as inflated singleton counts. When fitting the models above, we addressed these distortions in two ways. First, we excluded SNPs that are transitions, thus excluding false C->T mutations caused by deamination. Second, we did not ascertain SNPs within the samples with less than 9x coverage: MA1, Sidelkino, Okunevo, and ShamankaEN. We required all SNPs to be polymorphic when restricted to the higher-coverage samples, computing the SFS *conditional* on this ascertainment scheme. Note this automatically excludes all singletons within the low-coverage samples, since such SNPs would not be polymorphic within the higher-coverage samples.

In the ascertainment scheme above, SNPs were ascertained within Mbuti, Han, Yamnaya, KK1, Botai, and Sholpan. While Mbuti, Han, and Yamnaya are very high coverage (>20x), the Sholpan (9x), KK1 (11x), and Botai (14x) samples have modest coverage and are potentially susceptible to errors in

ascertainment. We thus reran our results, excluding these medium-coverage ancient samples from the ascertainment scheme. Our inferred demography is shown in Fig. S22, and it is nearly identical to the demography in Fig. S21. The biggest difference between the demographies in Figs. S21 and S22 is that the KK1-Yamnaya split time increases by a few thousand years, from ~20kya to ~24kya.

The similarity between Figs. S21 and S22 suggests the singleton counts for the medium coverage ancient samples are not distorted sufficiently to substantially change the outcome of the analysis, and that excluding the low-coverage samples (<9x) from ascertainment was sufficient to control for ascertainment error.

S2.14 Uniparental marker analysis

S2.14.1 Y-chromosome analysis

S2.14.1.1 Variant calling and haplogroup determination

We called Y-chromosomal variants in 45 ancient and 103 modern samples (Section S2.1) using bcftools (http://www.htslib.org/doc/bcftools.html) (*134*) mpileup and bcftools call emitting all sites within mappable Y-chromosomal regions (*135*). Haplogroup determination was done with the script callHaplogroups.py distributed with Yhaplo (*136*), with the parameter --ancDer, which outputs the allele counts for ancestral and derived SNPs along a path of branches of the Y-chromosome tree. In total, approximately 20,000 phylogenetically informative SNPs from the ISOGG 2016 database (http://isogg.org/tree/ISOGG_YDNA_SNP_Index.html) were used for haplogroup determination. Given the low coverage of the ancient DNA samples and the effect of deamination on lineage determination, we manually inspected ancestral and derived alleles to evaluate their phylogenetic consistency, ensuring that the lineages identified were the most likely considering the data observed. Y-chromosome lineages are presented in Table S13 as well as ancestral and derived counts in Table S14.

S2.14.1.2 Y-chromosome phylogeny

We investigated Y-chromosomes in our dataset by first selecting 103 present-day individuals from Africa, Eurasia and the Americas, including the ones newly sequenced in the present work and 6 additional high-coverage ancient samples: Yamnaya (present study), Clovis (*137*), Ust-Ishim (*128*), Saqqaq (*138*), KK1 (*7*), and BR2 (*139*). We filtered heterozygous SNPs from this dataset to remove potential deamination and errors and selected variants with exactly 2 alleles, minimum allele count of 1, depth of coverage >=5 and genotyping rate 0.97, and restricted variants within callable regions of the Y-chromosome. This resulted in a VCF file with 19534 SNPs, which we converted to tab format with vcf-to-tab (*9*), and then to multi-fasta with vcf_tab_to_fasta_alignment.pl (http://code.google.com/p/vcf-tab-to-fasta). Next, we performed MUSCLE alignment (*140*) and built a maximum likelihood tree using MEGA7 (*141*), which we re-rooted on the African main clade A, to which 2 San and 1 Mbuti individuals belong.

S2.14.1.3 Adding low-coverage ancient branches to a tree estimated with high-coverage Y-chromosomal data

The ancient DNA (aDNA) field is abundant in low/medium-coverage data, but considering the difficulties inherent to estimating accurate phylogenies from it, datasets with large number of ancient samples are rarely represented in the form of a tree. Therefore, we aimed to incorporate low-coverage

ancient samples into a pre-computed Y-chromosome tree with high-coverage modern and ancient samples.

The main idea behind this approach is that haplogroup names identified in aDNA samples can be informative about their relative position on the tree. For example, a given sample carrying the M269-R1b1a2 lineage should be placed within the same clade as other R1b1a2-derived individuals. In the case where further downstream markers are not available for that particular sample, which would allow placing it at a more specific branch, the upper bound of confidence for placing the ancient sample would be at the node of all clade(s) containing R1b1a2-derived individuals. Based on this premise, we tried to map a set of Y-chromosome lineages identified in ancient samples to the most related lineages in a tree estimated with high-coverage data.

2 data structures are required: a tree estimated with high-coverage samples and a list of haplogroups identified in ancient low coverage individuals. First, we label each branch of the computed tree with the haplogroup identified for each one of the samples. Next, for each haplogroup in the list of ancient samples, we first attempt to identify matches in the lineages present in the tree. In the event that a single exact match is found, we replace that tip with a subclade containing the ancient sample lineage and the matching tip, as these samples are likely to form a clade in a Y-chromosome tree. In the case of multiple exact matches to the tips of the tree, the ancient sample is added to the node ancestral to those tips—i.e., the most common recent ancestor. In the case where no exact matches were found, we trim the query haplogroup identified in the ancient sample by 1 letter (for example, instead of searching for ‘R1b1a2a2’, we would now try to match ‘R1b1a2a’) and repeat the process, until one, or several partial matches have been identified. Given we are dealing with large amounts of missing data, we opted for the most conservative approach of binding ancient DNA samples to ancestral nodes containing all matches, than directly to the matching tips. The reason for this is simply because sequencing more data could reveal that a given sample belongs in reality to a more downstream branch of the tree. In this way, we only provide the upper bound of where we can confidently map ancient samples to the phylogeny. Using this procedure, we inserted 44 ancient DNA samples (40 from the present study and MA1 (*26*), Kennewick (*142*), Loschbour (*65*), and Bichon (*7*)) into a tree estimated with high-coverage sequences. Sample mapping to tree locations was confirmed by examination of ancestral and derived SNPs at the branches of the high coverage phylogeny.

In the cases where it was not possible to identify a fully resolved Y-chromosome lineage for a particular sample, the placement of ancient samples in a pre-existing phylogeny may still provide insights into population affinities and biogeographical distribution of ancient and modern haplogroups.

S2.14.1.4 Visualizing ancestral and derived SNPs

Given the incompleteness typical of low coverage ancient DNA data, full Y-chromosome haplotype resolution was not possible for the majority of our samples. With this in mind, we generated a visual representation of allele status and missing data at important branches of the tree for ancient and modern samples in our dataset.

Yhaplo’s default behavior uses a decision table, which specifies the number of ancestral and derived SNPs required to continue traversing the tree and which nodes to visit. In this mode, the output only includes derived and ancestral alleles observed in the tree path travelled for lineage assignment. We altered the code of Yhaplo so that positions with missing data (no alleles observed) were also outputted in addition to derived and ancestral alleles.

We used the table.4phylo function of the R package adephylo at each node to generate a table of allelic state at each branch of ISOGG Y-chromosome tree for each haplogroup. Next, we plotted the ISOGG Y-chromosome tree for the relevant nodes to which our ancient samples belong including the

aforementioned table with allele status information for each branch. Trees with allelic information for the N and Q clades to which the majority of our ancient samples were assigned are shown on Fig. S24.

S2.14.1.5 Limitations

We present an automated solution for incorporating low-coverage ancient samples into confident Y-chromosome phylogenies, which allows examining phylogenetic affinities with the available data. There are a few limitations inherent to our approach: first, when calling Y-chromosomal variants from low-coverage sequence data, not all lineage defining markers are covered by reads, and, therefore, aDNA samples may be positioned at more ancestral nodes in the tree when, in reality, more data could reveal that they may belong to a better resolved branch of the Y phylogeny. A second limitation of this method is that it only uses known markers which were ascertained in modern populations to determine membership to Y-chromosome lineages, and, therefore, unknown variants are not being used to place low-coverage samples onto the tree. Furthermore, this method depends on haplogroup nomenclature, which may change as more SNPs get discovered and as the nomenclature system is updated. Lastly, by adding branches to the tree on the basis of haplogroup name and not by estimating genetic distance results in loss of branch length information. With this in mind, we urge caution interpreting phylogenetic affinities estimated with low-coverage aDNA samples, due to known problems such as deamination and incompleteness of the data.

S2.14.1.6 Results

S2.14.1.6.1 Steppe – Botai and Yamnaya

We identified 2 distinct Y-chromosome lineages in the two Botai_CA male samples: BOT14 was determined to carry a derived allele at M478-R1b1a1 and BOT15 belonged to the basal haplogroup N. In the phylogenetic tree (Fig. 5), the R1b1a1 sample BOT14 is paired with a single individual from the Teleut population of southwestern Siberia/Altai. The marker M478 belongs to the same branch as M73 and both define the R1b1a1 lineage, which occurs almost exclusively in non-Europeans (*34*). This lineage reaches the highest frequencies in Central Asia and Siberia, in particular in populations surrounding the Altai region, such as the Kumandins (35%) (*143*), Bashkirs (23%), and Balkars (10%) (*34*).

The newly sequenced high-coverage Yamnaya sample carries the R1b1a2a2c1 lineage, which is closely related to R1b1a2a2 previously identified in other Yamnaya samples (*2*) and can be commonly found in present-day Eastern Europeans and in the Caucasus region. In the phylogenetic tree, this sample was placed more closely to one R1b1a2a2 Avar and 1 Okunevo individual. The Upper Paleolithic MA1, whose ancestry is present in both Yamnaya and Botai, carries derived alleles at markers defining the basal R haplogroup, and, therefore, it is placed at the root of all R clades. The geographical distribution of R clades found in our dataset can be seen in Fig. S25.

S2.14.1.6.2 Baikal Early Neolithic

In the Baikal_EN males, N subclades occur in all samples, except for DA250, which belongs to NO1. However, more data may reveal membership to a more downstream clade of the Y-chromosome tree. We have determined Ust-Ishim to belong to a more ancestral lineage ancestral NO lineage, in agreement with recent re-examinations of this sample's Y-chromosomal affinities (*115, 136, 144*). Also in (*136*) the authors have pointed out that the Romanian Oase 1 sample (*145*) also shares this lineage,

which was probably widespread across Eurasia. The presence of subclades of haplogroup O in East Asia and N across Northern Eurasia is consistent with this hypothesis.

Of the remaining samples, individual DA247 belongs to the N lineage and DA251 to N1 but with no possibility of determining N1c2 due to the lack of reads covering the defining markers of this lineage. In our phylogenetic tree, DA245, DA248, and DA362 form a clade with 1 Komi individual and 1 Khanty individual, which all belong to N1c2 (Fig. S26). We note that we have excluded marker L665 that determines N1c2b2, given it presented clear inconsistencies with the haplogroup affiliation of some of the samples, with some presenting the derived allele at L665, but the ancestral allele for many markers upstream of this marker. Sample DA357 presented derived alleles at markers defining C2b (1 ancestral, 3 derived), C2b1 (2 derived), and C2b1a1 (1 derived), which points to a likely assignment to C2b1a1. However, it is worth noting an ancestral allele at C and a derived allele at N1c2, which bring uncertainty to haplogroup determination. Nonetheless, an N1c2 affiliation is unlikely because of an ancestral marker at N1 and considerable support for this sample to belong to C2b1a1.

S2.14.1.6.3 Late Neolithic/Bronze Age Baikal and Okunevo

After the Early Neolithic, the archaeological record of the region surrounding the Baikal Lake is characterized by the absence of burial sites that only reappear 1,500 years later during the Late Neolithic (*88*). After that, the Bronze Age cultures emerge in the area. It was therefore interesting to determine whether there were genetic shifts accompanying these cultural transitions. Additionally, PCR-based studies of these remains had already strongly suggested the presence of discontinuity between the EN and LN/BA at the level of Y-chromosomes (*89*).

As observed in Fig. 5, the transition observed between the Early Neolithic and Early Bronze was characterized by complete Y-chromosomal lineage turnover, with the former group carrying almost exclusively N lineages and the later presenting instead Q lineages. Interestingly, in the Okunevo culture from the Altai region, prevalence of Q lineages was also observed. It is worth noting that the lineages identified in 2 UstidaLN samples belong to both N and Q haplogroups: individual DA345 was classified as belonging to N1c1(xN1c1a), which has been reported to reach high frequencies (~80%) in the Yakuts (*146*). This sample was included in the same clade as other Siberian groups, such as Buryats, Yakuts, and Bashkirs. However, due to missing data, it was not possible to discern if this sample is ancestral to all these individuals or instead can be grouped with a particular branch of the tree. The other UstIdaLN DA355 carried a derived allele at M346, which defines Q1a2.

1 Okunevo sample and 1 Kurma sample were assigned to Q1a, but additional resolution was not possible given the sparsity of the data. One Okunevo sample (RISE683) belongs to Q1a1b1 (xQ1a1b1a), also identified in 1 Karasuk individual (*1*) and is extremely rare in present-day populations. In our modern dataset, 1 sample from Uzbekistan carrying Q1a1b1a is the closest match to Q1a1b1. We note that these lineages are distinct than the one presented by Saqqaq Q1a1a-F746, which is prevalent in Inuviats from the Canadian Western Territories (*143*).

The Okunevo individual RISE670 belongs to Q1a2b-L940 (xQ1a2b1,Q1a2b2), which has a mostly Central Asian distribution. In our modern dataset, 1 Dungan is the closest match.

2 Okunevo and 1 UstidaLN and UstidaBA individuals belong to Q1a2-M346. In (*147*) this lineage appeared only in 2 individuals, one from the South Asian Brahmin population and the other from European Croats. In our modern dataset, Q1a2 has been identified in a Tajik individual. However, given the incompleteness of allele state at informative positions, it is not possible to determine whether the majority of ancient samples indeed belong to Q1a2(xQ1a2a, Q1a2b), as the Tajik sample, or a further downstream marker defining Q1a2a or Q1a2b, and therefore they were placed at the root of all Q1a2 branches: DA355 Q1a2(xQ1a2b2,Q1a2a1b,Q1a2a1c); DA361 Q1a2

(xQ1a2b,Q1a2a1b,Q1a2a1a1,Q1a2a1a2); RISE672 Q1a2(xQ1a2a,Q1a2b1,Q1a2b2); and RISE674 Q1a2(xQ1a2a,Q1a2c,Q1a2b1,Q1a2b2).
In ancient groups, lineage Q1a2a-L53 was identified solely in the Baikal Early Bronze Age samples from Shamanka and Ust'Ida, which closely match one individual from Turkmenistan. Only individual DA336, which presents Q1a2a(xQ1a2a1), could be excluded from the downstream Q1a2a1 branch, with the others not having enough data to clarify their membership status. Despite this, the data obtained for a subset of Shamanka_EBA samples provided substantial evidence that these did not belong to either a Clovis-related branch Q1a2a1b defined by M971 or to Kennewick's M930-Q1a2a1a branch, specifically DA335 Q1a2a (xQ1a2a1a,Q1a2a1b2), DA337 Q1a2a (xQ1a2a1a,Q1a2a1c); DA338 Q1a2a (xQ1a2a1a,Q1a2a1b2); DA353 Q1a2a (xQ1a2a1a,Q1a2a1b,Q1a2a1c1); and DA356 Q1a2a (xQ1a2a1b,Q1a2a1a1d,Q1a2a1a1e).

1 Okunevo sample and 1 UstIda_EBA belong to Q1a2a1, and where data is available, these samples carry ancestral alleles at markers defining American lineages: DA343 Q1a2a1 (xQ1a2a1a,Q1a2a1b); RISE662 Q1a2a1 (xQ1a2a1b,Q1a2a1a1,Q1a2a1a2).

1 ShamankaBA (DA339) and 3 Okunevo (RISE664, RISE718, RISE719) belong to Q1a2a1c-L330 (xQ1a2a1c1), lineage also present in the Yeniseian-speaking Kets in our dataset. These lineages are also distinct from the ones identified in Clovis (Q1a2a1b-M971) and Kennewick (Q1a2a1a-M930). Geographical patterns illustrate well the regional differences in terms of Q lineages in our modern and ancient samples (Fig. S27): the Q lineages identified in our samples have a Central Asian/Siberian distribution, while the lineages identified in the Paleoamericans Clovis and Kennewick occur mostly in Native American populations.

Interestingly, 1 Okunevo individual (RISE675), presented the R1b1a2 lineage. However, by directly inspecting the BAM file we realized that by applying variant quality filters, these removed the derived allele A at the Z2105 marker (C->A), which defines the R1b1a2a2. This allele is indeed present in RISE675 although only covered by one read, supporting the notion of admixture with Yamnaya-related peoples (largely assigned to R1b1a2a2). In addition to this, the R1b1a1 lineage identified in Botai does not support a direct link between Botai and this Okunevo individual, though we urge caution interpreting these results given the small sample size of Botai males sampled in the present work (n = 2).

S2.14.1.6.4 Turkmenistan and Anatolia

The Namazga samples from Turkmenistan belong to J-M304 (DA379) and to J2a1-L26 (DA381). The later Iron Age sample Turkmenistan_IA from the same region belongs to the F992/Z93-R1a1a1b2 lineage, which has also been identified in Srubnaya Late Bronze Age Steppe (LBA) populations (*47*). In our dataset, this lineage and their subclades have been identified in 4 Altaians, 2 Kyrgyz, 2 Bashkirs, 2 Tajiks, 1 Teleut, and 1 Uyghur individual. In a larger survey of R1a derived males, it was determined that the vast majority of Z93 lineages occur in Central and South Asian groups, while the sister branch Z282 is mostly restricted to Central and Eastern Europe (*148*). The fact that the Turkmenistan_IA sample shares the Z93 lineage with Srubnaya is in agreement with the increased affinity of the Turkmenistan sample to LBA steppe populations.

All Anatolian Early and Middle Bronze Age individuals belong to J2a derived lineages with the exception of the Anatolian MLBA sample MA2208, which instead carries the G2a2b1 lineages, closely related to those present in Anatolian and European Neolithic samples (*47, 149*). Regarding the J2 lineages identified, transmission through contact with populations related to Caucasus hunter-gatherers or Iranian Neolithic groups is a possible explanation, given they have been shown to carry J/J2 clades (*7, 42*) (*150*).

S2.14.2 Mitochondrial DNA analysis

S2.14.2.1 Ancient sample mtDNA lineage determination

To investigate mitochondrial DNA lineages in our ancient and present-day dataset, we selected reads aligned to the mtDNA with samtools and uploaded the resulting individual bam files to the mtDNA server (*151*). We submitted the resulting hsd output file to haplogrep V2, which we used for haplogroup identification, and downloaded the resulting aligned mtDNA sequences in fasta format. The maximum likelihood phylogeny shown on Fig. 5 was generated with RAXML (*152*), GTRCAT model, and 100 bootstraps, selecting the best tree. In order to minimize uncertainty, we removed 3 samples whose position in the phylogeny did not match the haplogroup identified: Kurma DA354 (D4, haplogrep score 0.61), Anatolia_IA MA2197 (U8b1b2, 0.57), and Namazga_CA DA380 (U2b, 0.69).

S2.14.2.2 Results

We identified a diverse set of mtDNA lineages in our ancient samples belonging to the main clades A, C, D, F, G, H, J, K, R, T, U, W, and Z (Table S15).

Regarding lineage A, 7 Okunevo individuals were included in the A8a (n = 4) and A8a1 (n = 3) clades. A8 mitochondrial lineages are widespread in Far Eastern and Northern Siberian populations, such as the Dolgans, Itelmens, Evens, Koryaks, and Yakuts (*153*), and in our present-day data it has been detected in 1 Koryak individual. Additional distinct subclades of A were identified in 1 Lokomotiv (A), 2 UstIda_LN (A, A2), and 2 additional Okunevo (A) samples. Of these, the A2 lineage present in one UstIda_LN sample is of particular interest, given its subclades occur especially in Chukchis, Eskimos, and Na-Dene-speaking peoples (*153*). In the present-day dataset we analyze here, it has been found in individuals of the Yukpa, Tsimshian, Athabaskan, and Mayan populations.

The C5c lineage was identified in 4 Okunevo individuals. Interestingly, this lineage has been suggested to be restricted to Altai populations, which would suggest some extent of temporal mtDNA continuity in the region where Okunevo samples were excavated (*143*). We identified the C4a2a1 in 3 ShamankaBA, 1 Kazakh individual as well as closely matching 1 Yakut (C4a2a1a) and 1 Evenk individual (C4a2a1b). 2 UstIdaBA and 1 ShamankaBA carry the mitochondrial lineage C4a1a3, which was also identified in 1 ancient individual from Ust'-Belaya, dated between 4410–4100 BCE (*154*). CentralSteppe_EMBA samples both present subclades of the C4 lineage, with one of the samples carrying C4 and the other C4a1a4a. Regarding modern samples, our results are concordant with other observations that have shown that while C4a1 lineages are more widespread across Siberia, C4a2 are more restricted to Evenks and Yakuts (*155*).

The Copper Age Botai sample BOT2016 is placed as the root of the Z clade, and it presents haplogroup Z1a. In our modern dataset, haplogroup Z1a was found in an Altay-speaking Teleut individual and it has been reported to be broadly distributed across East/Central Siberia (*156*). Notably, the presence of the Z1a lineage in Saami, Finns, and Volga peoples has been linked to movements from Siberia into Northern Europe occurring between 3,000–2,000 years ago (*157*).

Clade D appears to have persisted in the Baikal region from the Neolithic to the Early Bronze Age, with occurrences of lineage D4 lineages across this period of time. Of these, lineage D4e1 occurs exclusively in 2 ShamankaEN. The mitochondrial lineage D4j, however, was identified in both Baikal Neolithic and Bronze Age individuals and typically presents a South Siberian distribution (*158*). Additionally, D4j was also found in 1 Ottoman individual, which may be the result of contact with Central Asia during this period, as also supported by autosomal ancestry observations for this sample. In our modern dataset, multiple subclades of D4 were identified in Dai, Buryat, Teleut, and Khanty

individuals. We note that the North American Clovis sample carries the D4h3a7 haplotype and that a Devil's Gate Neolithic sample belongs to the D4 haplogroup (*25*).

In the ancient samples of the present study, clade G is represented by 3 ShamankaEN and 1 ShamankaBA individuals that belonged to G2a1, a subclade of G2a that is mostly restricted to Central Asian populations (*159*). Interestingly, we note one Scythian individual presented a closely related haplogroup, G2a4 (*47*). G2a is frequent in Turkic- and Mongolic-speaking populations in Asia (*158*), which is in agreement with the higher amount of Eeast-Asian-related ancestry identified in the Baikal Neolithic group. In the present-day dataset, it is more closely related to 1 Buryat and 1 Dungan, which present subclades of G2a and G2b, respectively.

Regarding clade H, 3 Okunevo individuals belong to H6a1b and one to H6a. The closely related H6b was also identified in one Tajik individual. H6 lineages can be commonly found in Central Asian populations (*158*).

Lineage F1b and sublineages were identified in 3 Baikal_EBA and 1 Baikal_EN, and in 1 individual each of the present-day Kalmyk, Turkmenistan, and Kyrgyzstan populations. F1b lineages have been reported in two 15-19th century Yakut individuals (*160*).

One other Botai sample (BOT15) presents the R1b1 lineage, which is also shared by an UstIda LN sample. Curiously this lineage has also been identified in a WHG (*139*). Yamnaya belongs to haplogroup R1a1a, and, interestingly, it has been found in peoples of the Caucasus and Eastern Europe, which is in agreement with the CHG and EHG composition of this archaeological group.

Regarding haplogroup K, it was identified in a Botai Copper Age sample (BOT14) that carried the mitochondrial lineage K1b2, with closest match in 1 Kazakh individual (K1b2a2). 2 samples from Anatolia also belonged to K, of which 1 Anatolia_MLBA sample presented the K1a haplogroup, present in both Europe and the Near East, and 1 Anatolia_Ottoman to haplogroup K.

Regarding clade U, it was identified in MA1(*26*), Sidelkino EHG (U5a2), and 1 Anatolia_MLBA (U1a).

The majority of Anatolian Bronze Age samples belong to J derived lineages (J2b1, J1c10a, J1c), and 1 Namazga sample from Turkmenistan carried J1. J2b is typically found in Atlantic and Mediterranean Europe, and J1c is widespread in Europe and commonly found in Neolithic remains (*161*). Lastly, Turkmenistan_IA DA382 was assigned to T2c1a, with a hypothesized Middle Eastern origin (*161*).

S2.15 Rare variant sharing between modern populations and the Botai and Yamnaya samples

Arjun Biddanda, Rui Martiniano & John Novembre

To further understand the distinct histories of Yamnaya- and Botai-associated ancestry in Eurasia, we carried out an analysis of rare-variant sharing. This analysis leverages the availability of whole-genome sequences for each sample and the whole-genome reference panels provided by the 1000 Genomes (1000G Project Consortium, 215) and Simons Genome Diversity Projects (*162*). Rare variants are typically the result of recent mutations that have taken place since the out-of-Africa dispersal and are geographically distributed in patterns that reflect the dispersal of descendants from the original carrier of the mutation (*163*). As such, they can provide useful markers of dispersal and recent ancestry (*164*).

S2.15.1 Relative abundance of rare variant sharing with European and East Asian populations at a regional scale

We first merged the dataset consisting of ancient whole-genome sequences from the Botai, Yamnaya, and other samples across Eurasia with the individuals from the 1000 Genomes Project. We then removed all variants that were either C-to-T or G-to-A transitions to avoid confounding due to DNA damage (*165*). This merged dataset was used to assess rare variants shared between ancient genomes and modern populations. We determined rare variants to be variants that had a global minor allele frequency < 1% in the 1000 Genomes Project Phase 3 dataset.

To explore broad-scale spatial patterns of rare variant sharing between ancient and modern genomes, we determined the number of rare variants that were shared between European populations (EUR) and East Asian populations (EAS) of the 1000 Genomes Project for each of several ancient sequenced genomes (Fig. S33). The Yamnaya consistently share a higher proportion of rare variants with European populations, whereas the Botai share a higher proportion of rare variants with East Asian populations (Fig. S33).

S2.15.2 Contemporary geographical distribution of rare variants that are shared with Yamnaya and with Botai

As a more fine-grained assessment of rare-variant sharing, we next sought to reveal the geographic distributions of contemporary rare variants that are shared with Yamnaya and with Botai. We took an approach that first involves categorizing variants by their geographic distributions. For each variant we then created a vector of length 26 where each entry in this vector represents the frequency of the variant in each of the 26 populations from the 1000 Genomes project. We then assemble all variants into a matrix and applied hierarchical clustering with K = 20 on the SNP-by-SNP distance matrix computed using the Canberra distance (*166*). For clustering we use the partitioning-around-medoids (PAM) with the cluster library for the R statistical software (*167*). The resulting categorical assignments and the frequency of variants that fall in each category allow for visualization of rare variant sharing patterns (Figure S34). We also compare the abundance of each category between Yamnaya- and Botai-shared variants, and we see that the Botai show a higher abundance of variants that are found exclusively in East Asian and American populations (Fig. S34).

S2.15.3 Geographic maps of rare-variant sharing abundance

As a second, more fine-grained, approach to assess rare-variant sharing approaches, we merged the ancient whole genome sequences with the Simons Genome Diversity Project (SGDP) (*162*) data, due to their finer scale sampling across the globe. Here we used the same set of variants that were rare (MAF < 1%) in the 1000 Genomes and counted the number of these variants that were shared between individuals in the SGDP and each ancient genome. We then plotted maps of the number of rare variants that were shared (Fig. S35). From Fig. S35 we see that Botai have a higher number of rare variants shared with individuals at higher latitudes and among Siberian populations, whereas Yamnaya share much more with European and South Asian populations.

S3: Radiocarbon dating

Karl-Göran Sjögren

4 human tooth and petrous bone samples from Kara-Depe, Geoksyur, and Takhirbai 3 were dated at the Chrono Centre, Queens University, Belfast. A further sample from Takhirbai 3 failed due to poor collagen preservation. Collagen extraction and other laboratory methods used at the Chrono Centre are described in detail in (*168*). Details of the datings are given in Table S3. 14C values were calibrated to 2 sigma intervals at the Belfast laboratory using the Calib software, rev 7.0.0, and the Intcal13 calibration curve. δ13C and δ15N were measured on all samples, as well as C/N ratio. C/N for all samples was between the accepted standard for good collagen quality, i.e. between 2.9 and 3.6.

The calibrated values in Table S3 do not take account of possible reservoir effects. The δ13C values of the samples are within the range for populations subsisting mainly on a terrestrial C3 diet, although slightly higher than usual. If C4 plants were also consumed, this would probably have been only in minor quantities. The δ15N values on the other hand, are higher than expected from such a diet. This may be due to several factors. First, the location of the sites in the vicinity of rivers suggests the possibility of a freshwater fish component in the diet, and the dates may in this case be affected by a freshwater reservoir effect (FRE). Second, elevated δ15N values may result from environmental factors such as dry climate and/or elevation. Third, since the analyzed samples consist of tooth and petrous bone samples, it is possible that the δ15N values are affected by a breastfeeding effect.

It is difficult to evaluate these possibilities on the basis of available data. Freshwater reservoir effects have not been studied in the region, and data from faunal remains at settlements are also not available. The extent of fish consumption is therefore unknown. The present climate of Turkmenistan is indeed arid, and much of the country is occupied by the Karakum desert. The locations of the studied sites at the foothills of the mountains in the south are characterized by slightly higher humidity than areas further north, but it is still arid. It is therefore quite possible that δ15N might be elevated due to climate. Regarding a possible lactation effect, the 2 sampled teeth were not determined, so we do not know which teeth were analyzed.

There is a possibility that the dates may be affected by an FRE of unknown size, although factors such as climate and lactation may well be sufficient to account for the high δ15N values. Also, the correspondence of the Kara-Depe dates with the commonly accepted datings for Namazga III suggests that the FRE may not be exceedingly large.

Fig. S1.
A plan of the excavation illustrates the burials of the skeletons (Namazga).

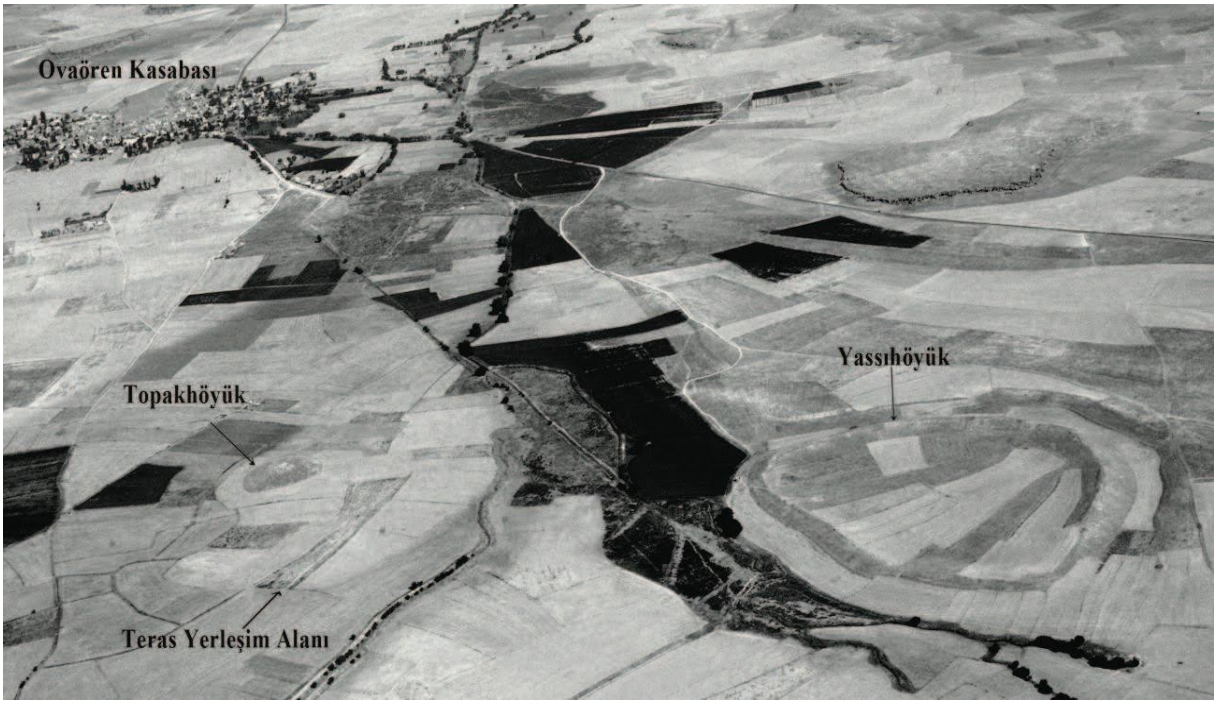

Fig. S2.
The three areas of Ovaören – Yassıhöyük, Topakhöyük, and the Terrace (Teras).

Fig. S3.
Graves and the well of trench GT-137, layer V.

Fig. S4.
Skeletons in the well of trench GT-137, layer V.

Fig. S5.
Geographical location of 41 newly sequenced present-day high-coverage genomes.

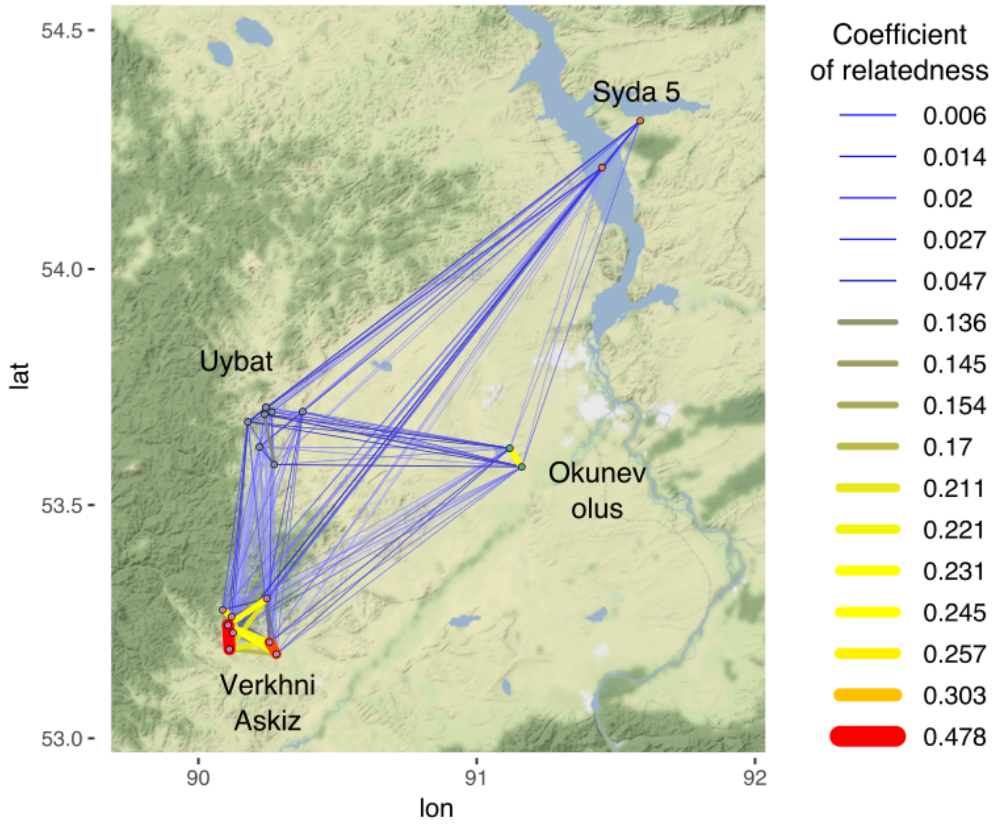

Fig. S6.

Geographic patterns of relatedness between different Okunevo_EMBA groups from the Altai region (Verkhni Askiz, Okunev Ulus, Uybat and Syda 5) and between individuals within each group. Dots represent individuals, and the lines connecting them are colored according to the relatedness shared by those individuals. Coordinates were jittered slightly to avoid overlap between samples.

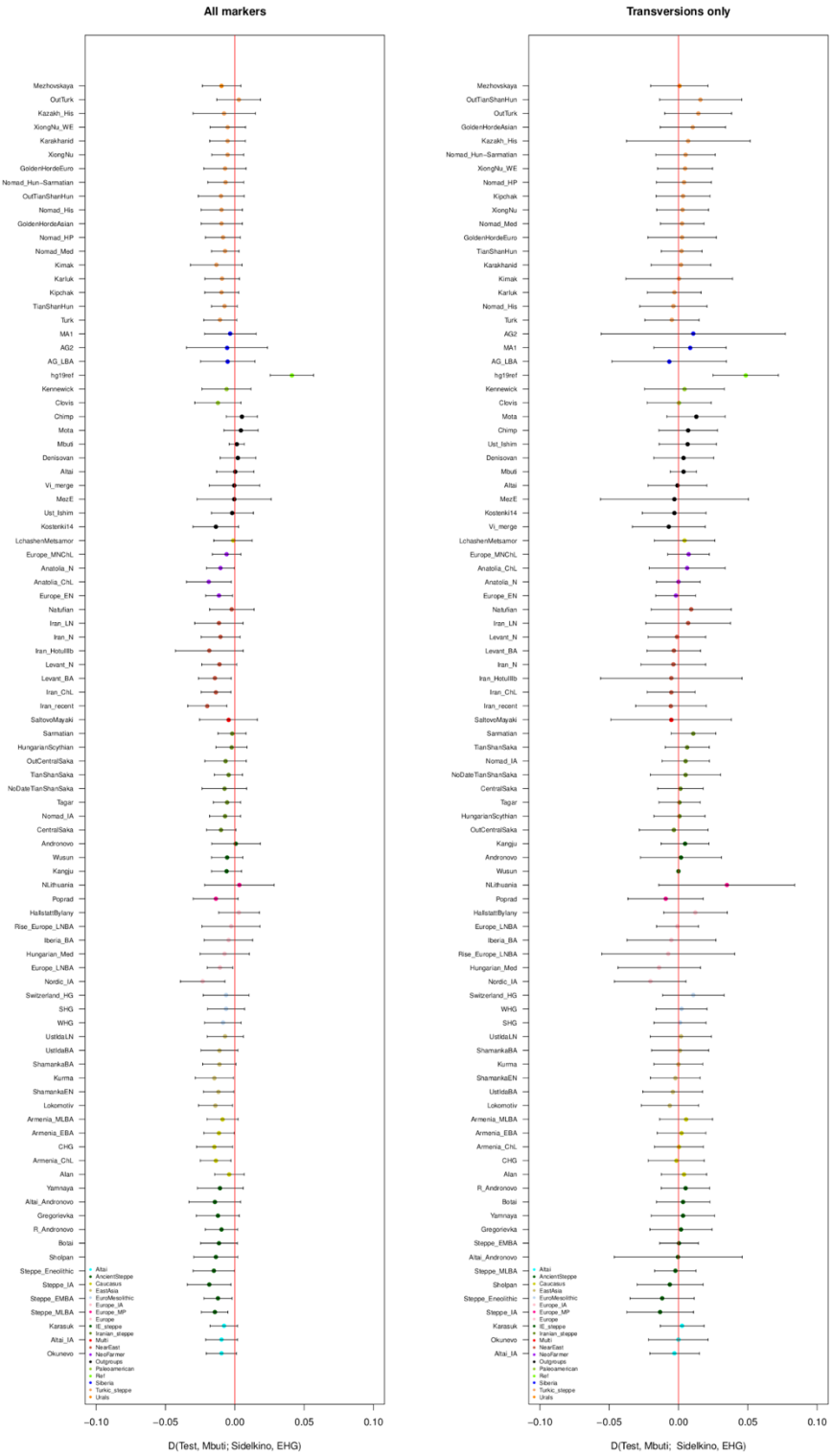

Fig. S7.

D-statistics test of the form D(Test, Mbuti; Sidelkino, EHG).

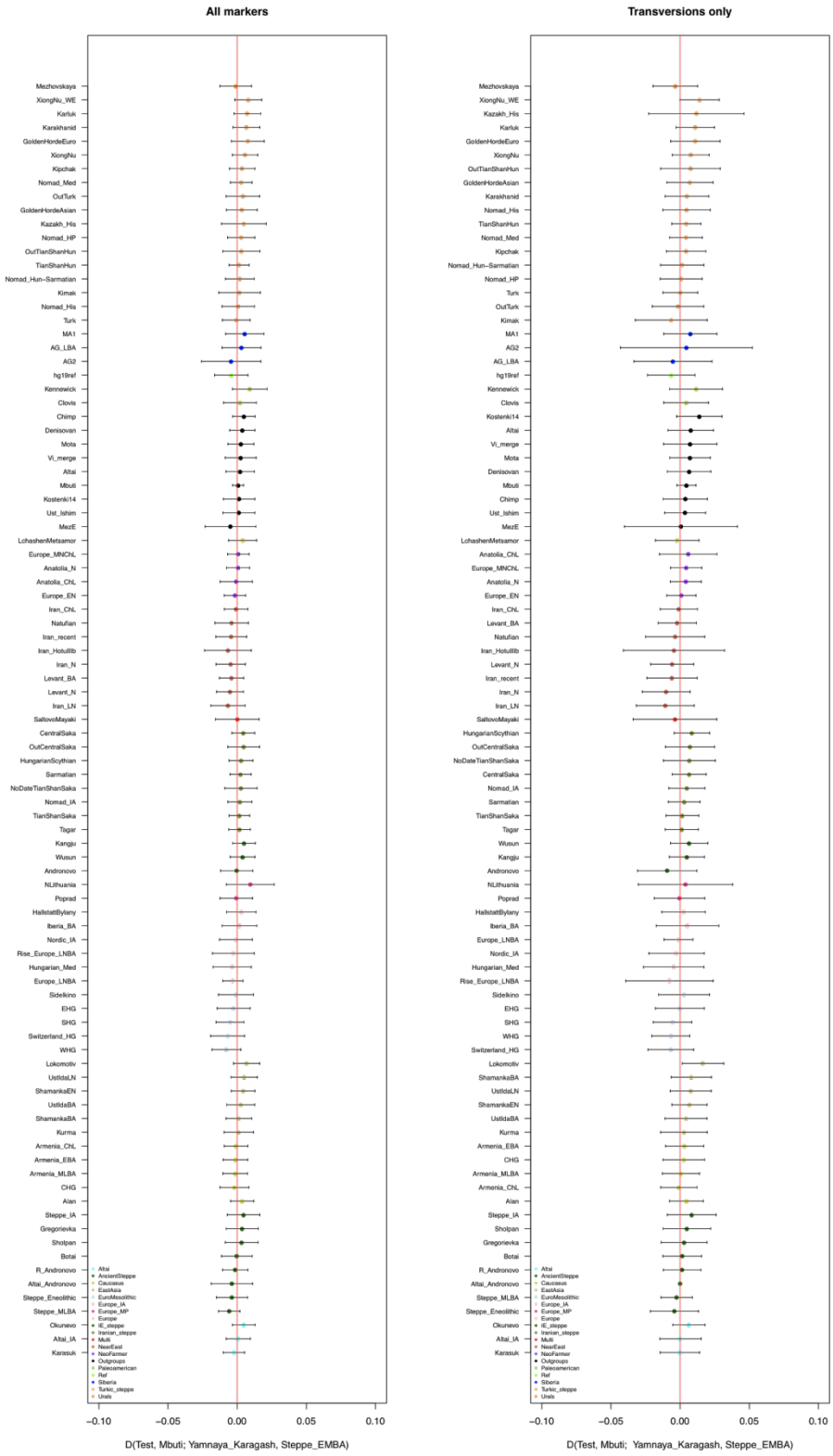

Fig. S8.

D-statistics test of the form D(Test, Mbuti; Yamnaya, Steppe_EMBA).

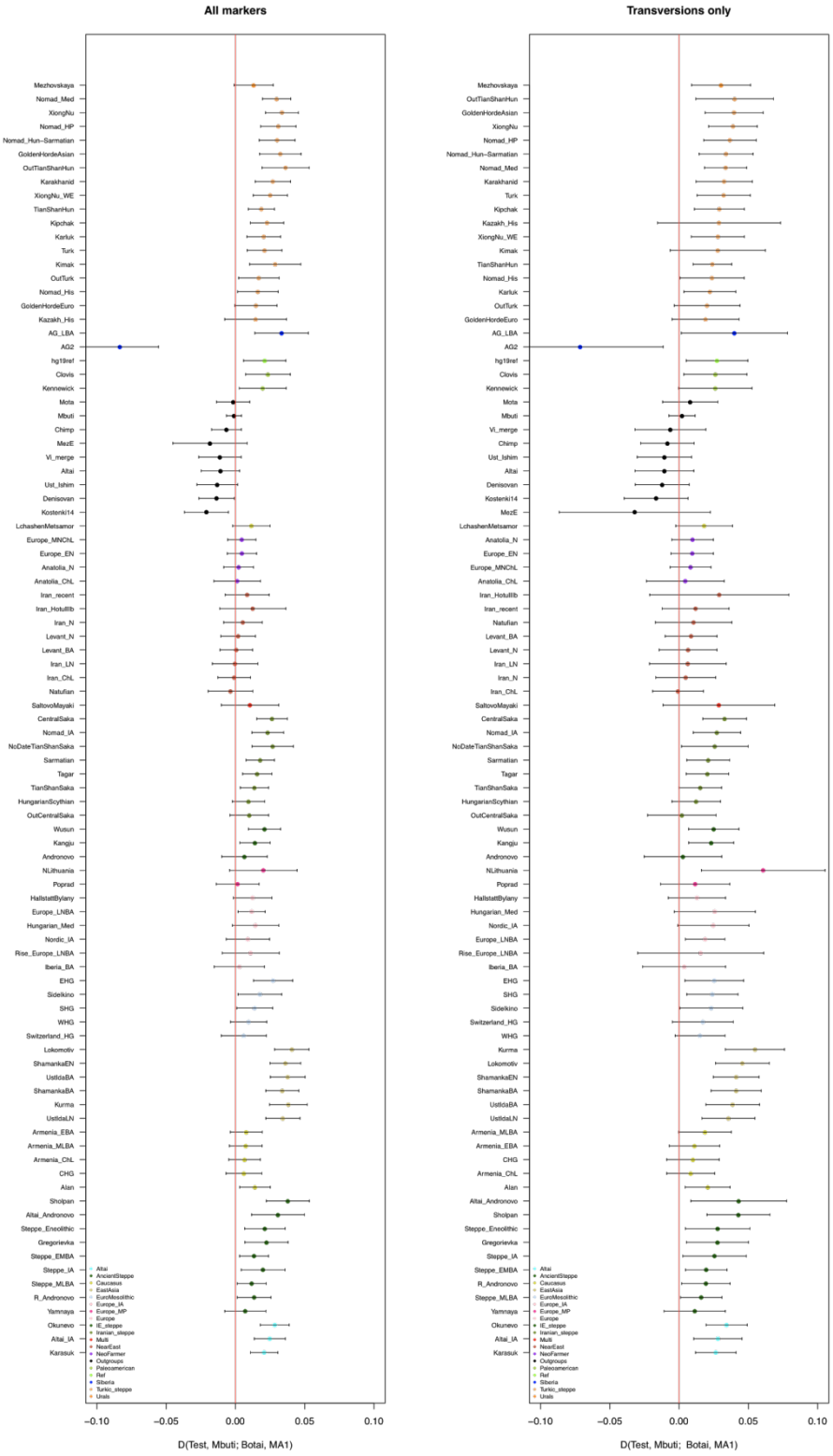

Fig. S9.

D-statistics test of the form D(Test, Mbuti; Botai, MA1).

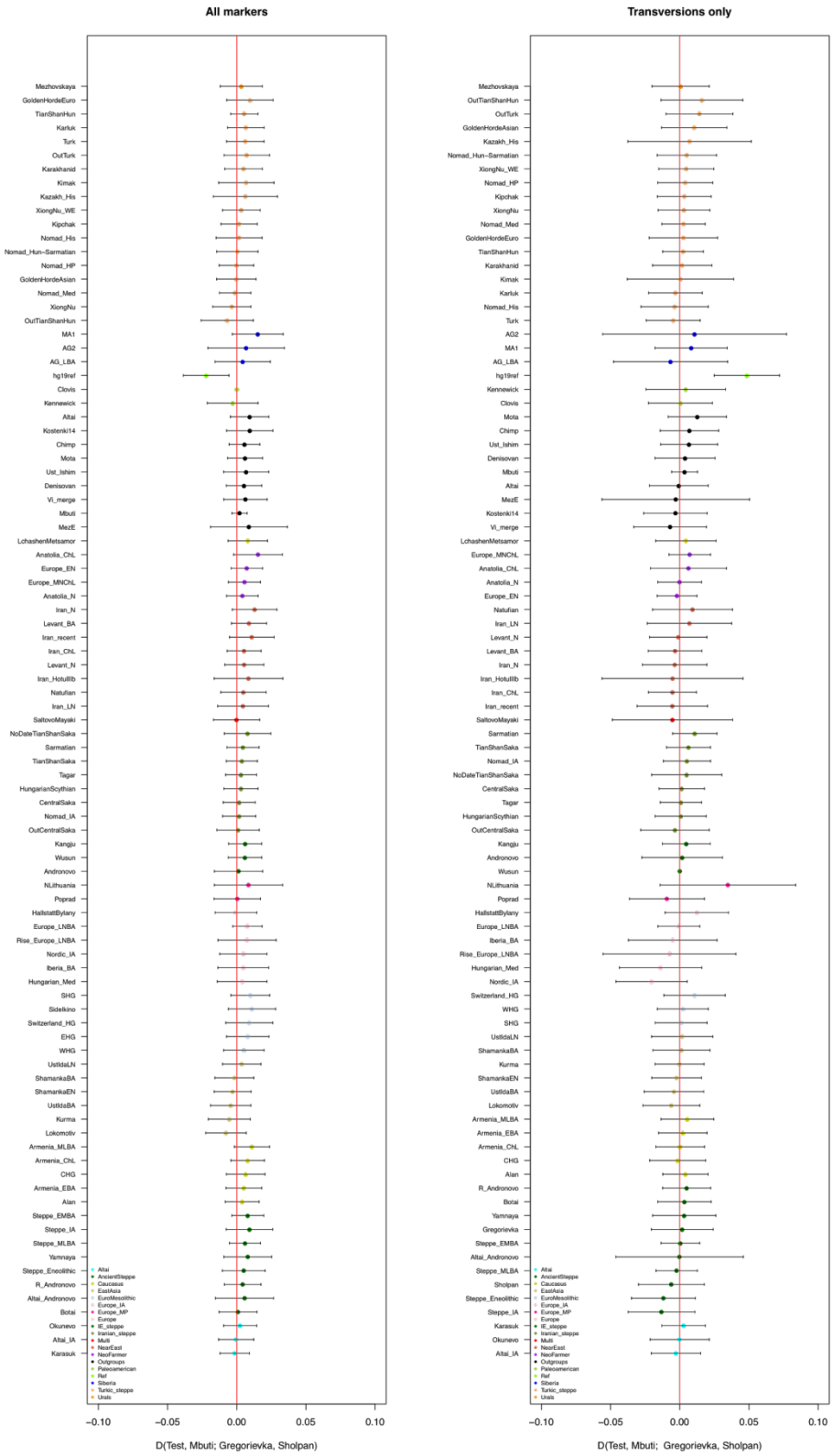

Fig. S10.

D-statistics test of the form D(Test, Mbuti; Sholpan, Gregorievka).

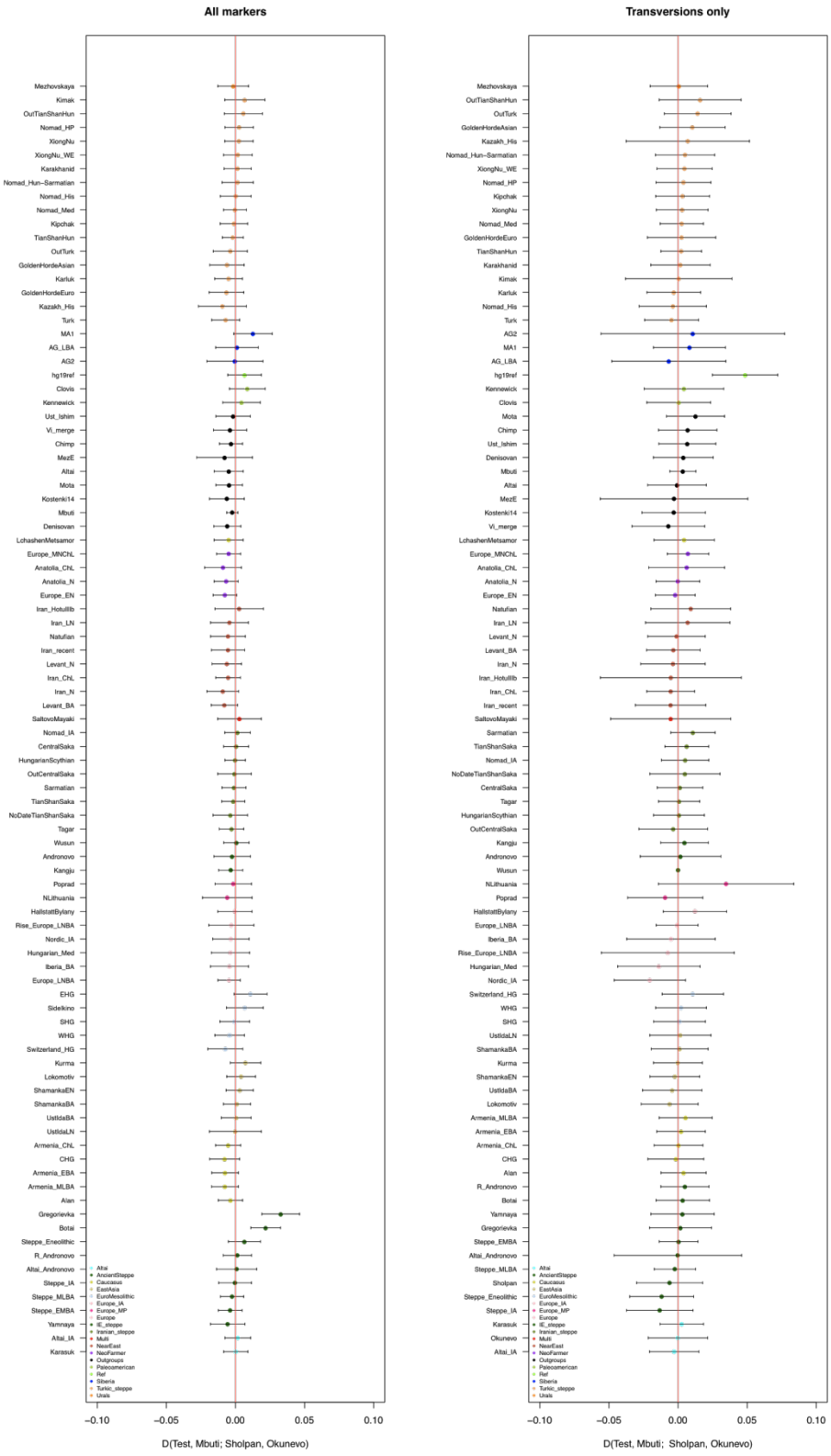

Fig. S11.

D-statistics test of the form D(Test, Mbuti; Sholpan, Okunevo).

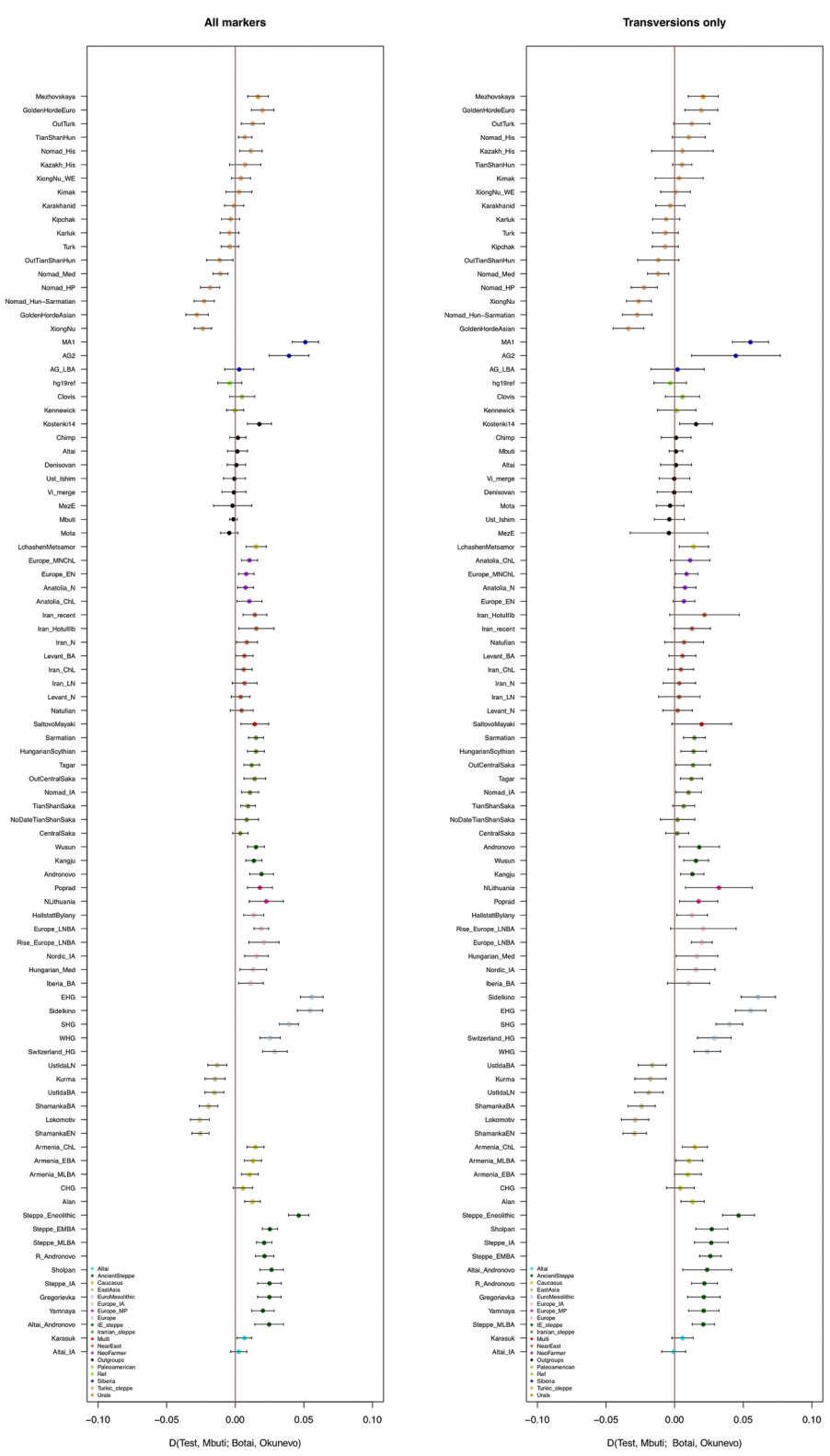

Fig. S12. D-statistics test of the form D(Test, Mbuti; Botai, Okunevo).

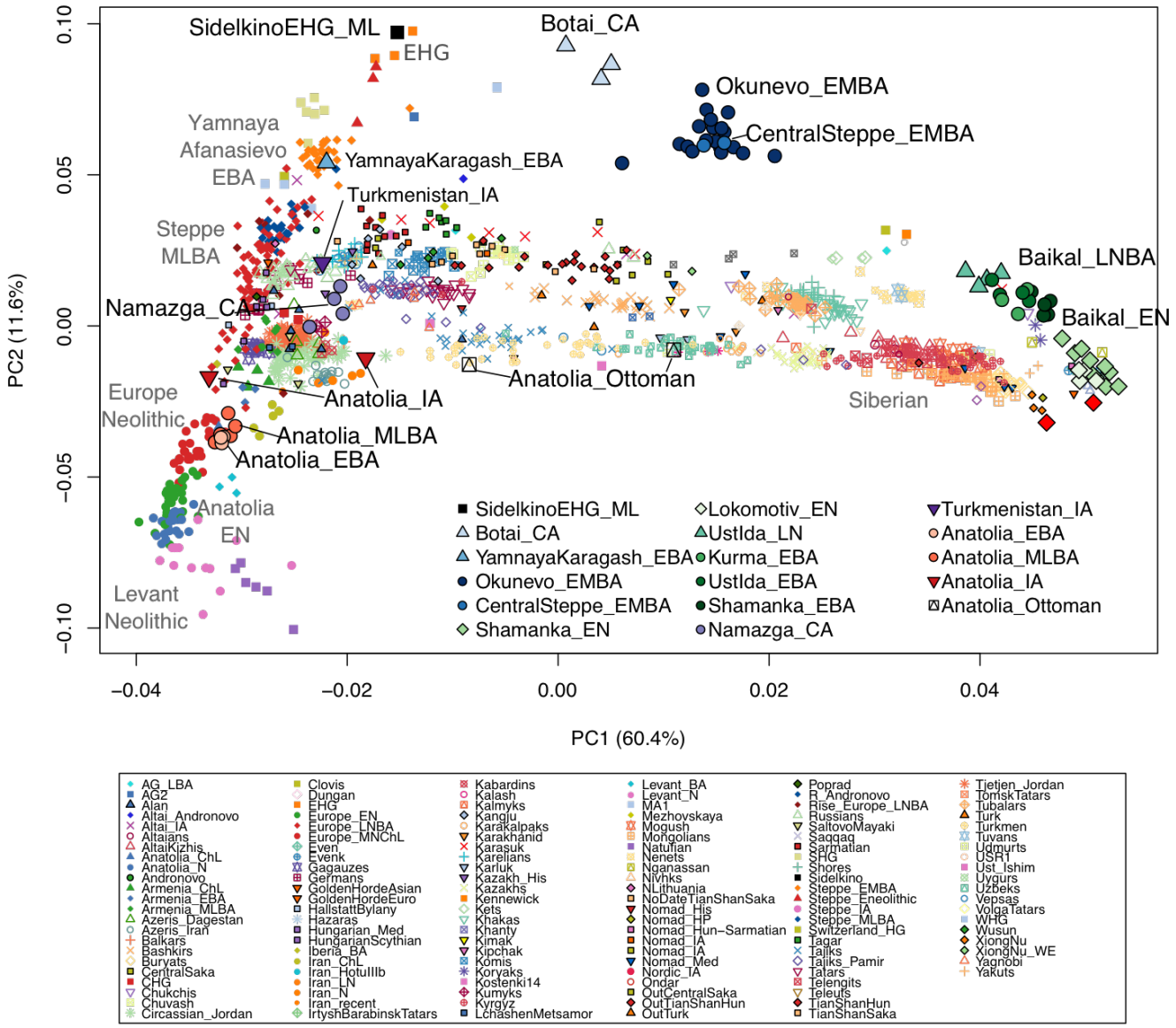

Fig. S13.

Principal Component Analysis estimated with ancient and modern Eurasians.

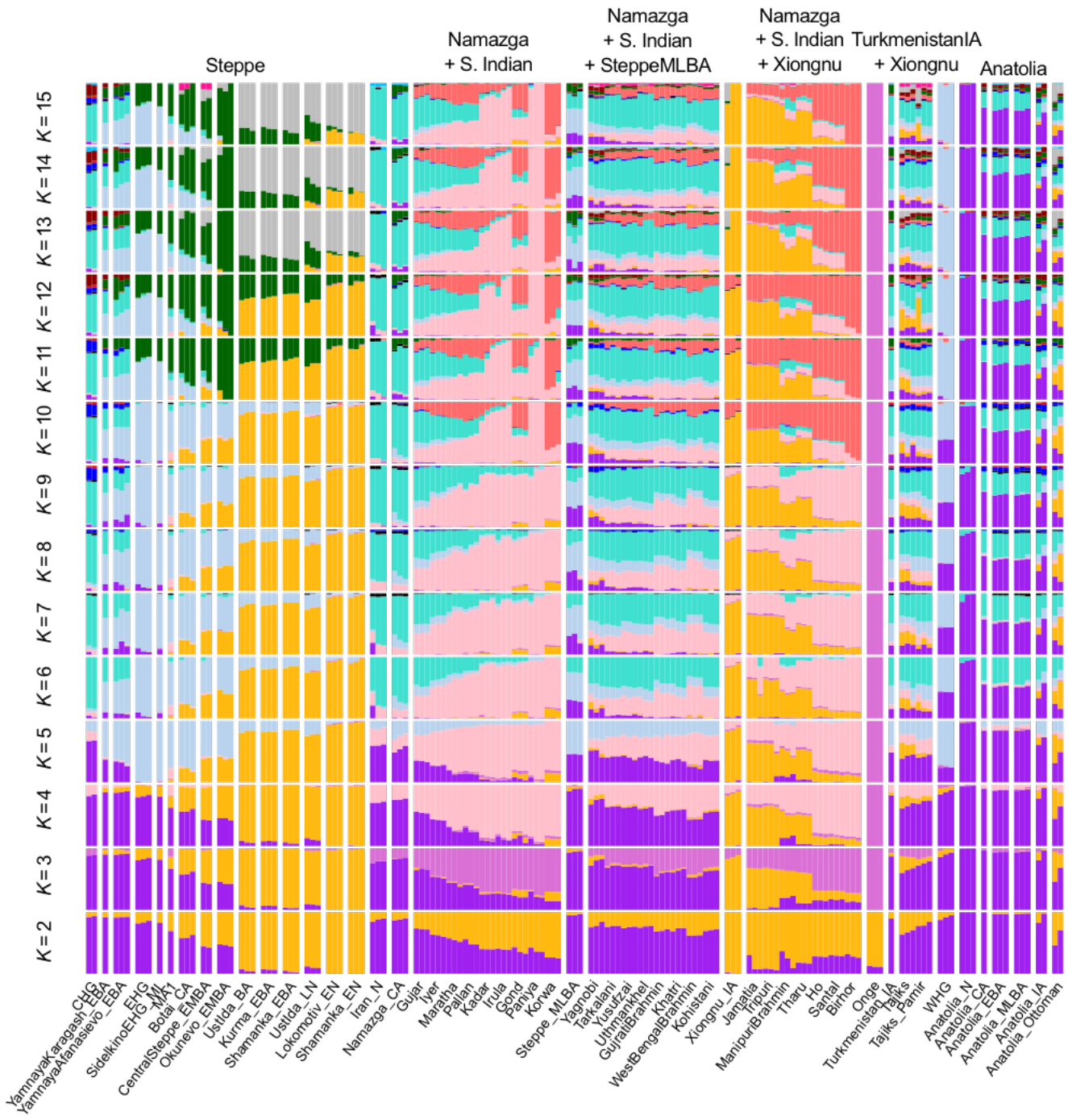

Fig. S14.

ADMIXTURE analysis for K = 2–15.

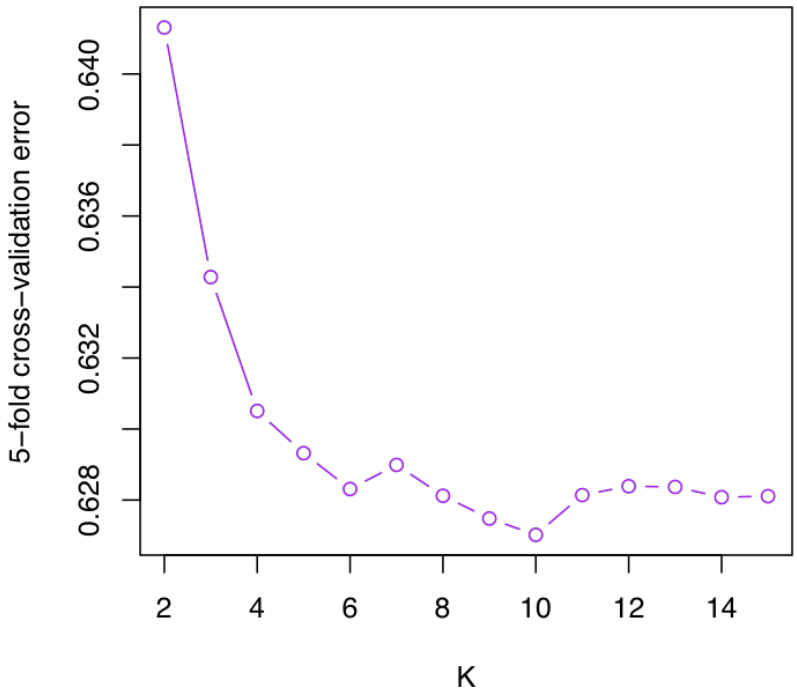

Fig. S15.
Cross-validation errors for the ADMIXTURE analysis.

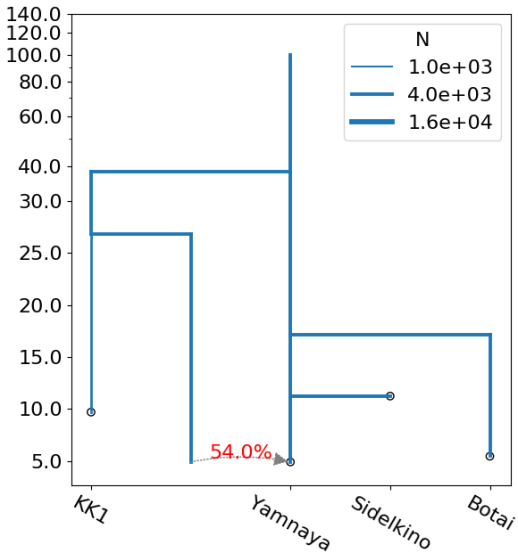

Fig. S16.

A simple 4-leaf demography centred around Yamnaya. Yamnaya was modeled as a mixture of CHG- and ANE-related ancestry, with 54% of its ancestry inferred to come from CHG. The Yamnaya ANE ancestry is inferred to be closer to Sidelkino (diverging 11 kya) than to Botai (diverging 17 kya). The Yamnaya CHG ancestry is inferred to be distantly related to KK1, diverging 27 kya, though we note the larger models in Figs. S19 and S20 inferred a more recent divergence time.

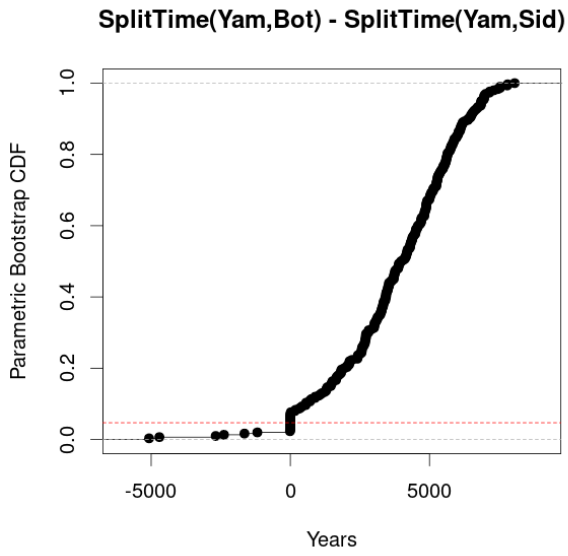

Fig. S17.

Parametric bootstrap distribution of $T_{\text{Botai-YamANE}} - T_{\text{Sid-YamANE}}$. The hypothesis $\{T_{\text{Botai-YamANE}} < T_{\text{Sid-YamANE}}\}$ can be rejected with $p = .047$ (shown in the red line).

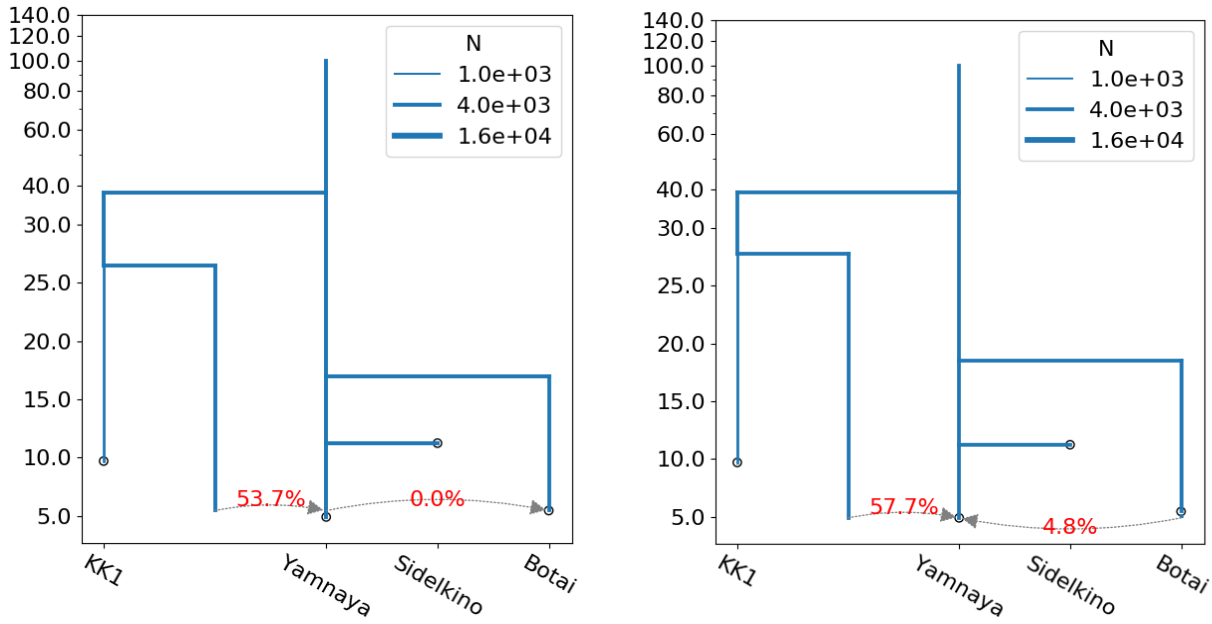

Fig. S18.

Adding additional gene flow events to the model in Fig. S16, we inferred no gene flow from Yamnaya to Botai, and a pulse of 4.8% from Botai to Yamnaya, which was not significantly different from 0 (p-value .18) under 300 parametric bootstraps simulated under the null model with no admixture (Fig. S17).

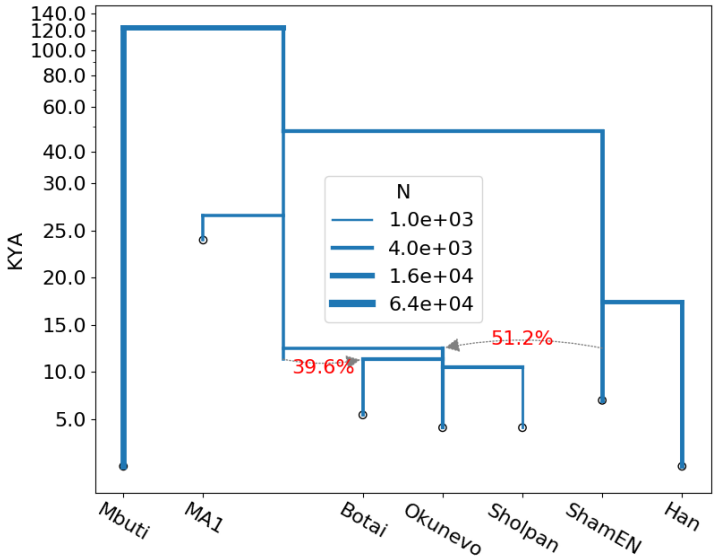

Fig. S19.

An inferred demographic model with 7 leafs, including 3 ancient steppe populations (Botai, Okunevo, and Sholpan) and 1 Baikal population (ShamankaEN). The steppe populations are modeled as a mixture of ANE ancestry (related to MA1) and East Asian ancestry (related to ShamankaEN). Botai has less East Asian ancestry than Okunevo and Sholpan, which we modeled by an additional ANE pulse into Botai from a ghost ANE population.

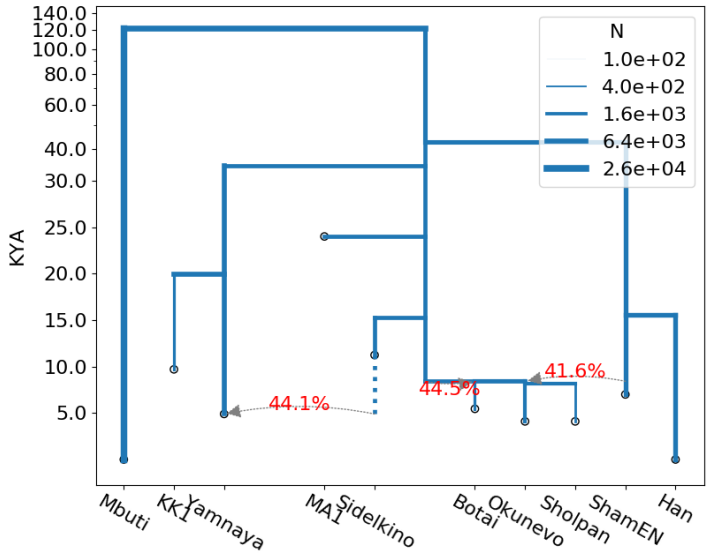

Fig. S20.

A 10-leaf model based on combining the models in Fig. S16 and Fig. S19 and re-estimating the model parameters.

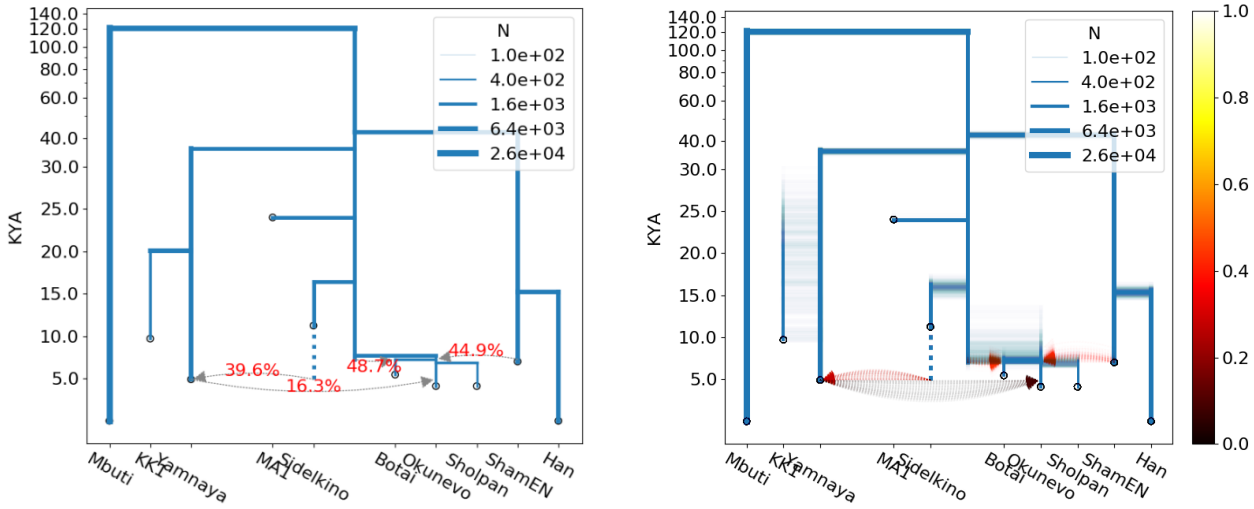

Fig. S21.

Our final estimated model, obtained by adding a Yamnaya->Okunevo pulse to Fig. S20. This is the same as the demography shown in Fig. 4 of the main text. On the left is our final point estimate; on the right we show 300 parametric bootstrap simulations, overlaid with transparency.

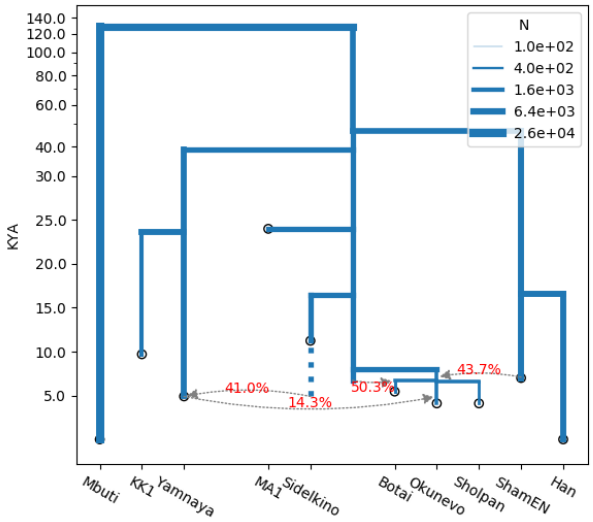

Fig. S22.

The result from refitting the demography in Fig. S21 but excluding Botai, Sholpan, and KK1 from ascertainment (along with Sidelkino, MA1, Okunevo, and ShamEN), so all SNPs are ascertained on the very high-coverage (>20x) Yamnaya, Mbuti, and Han samples. The inferred result is nearly identical, suggesting that potential errors such as inflated singleton counts in Botai, KK1, and Sholpan are not substantially biasing the inference.

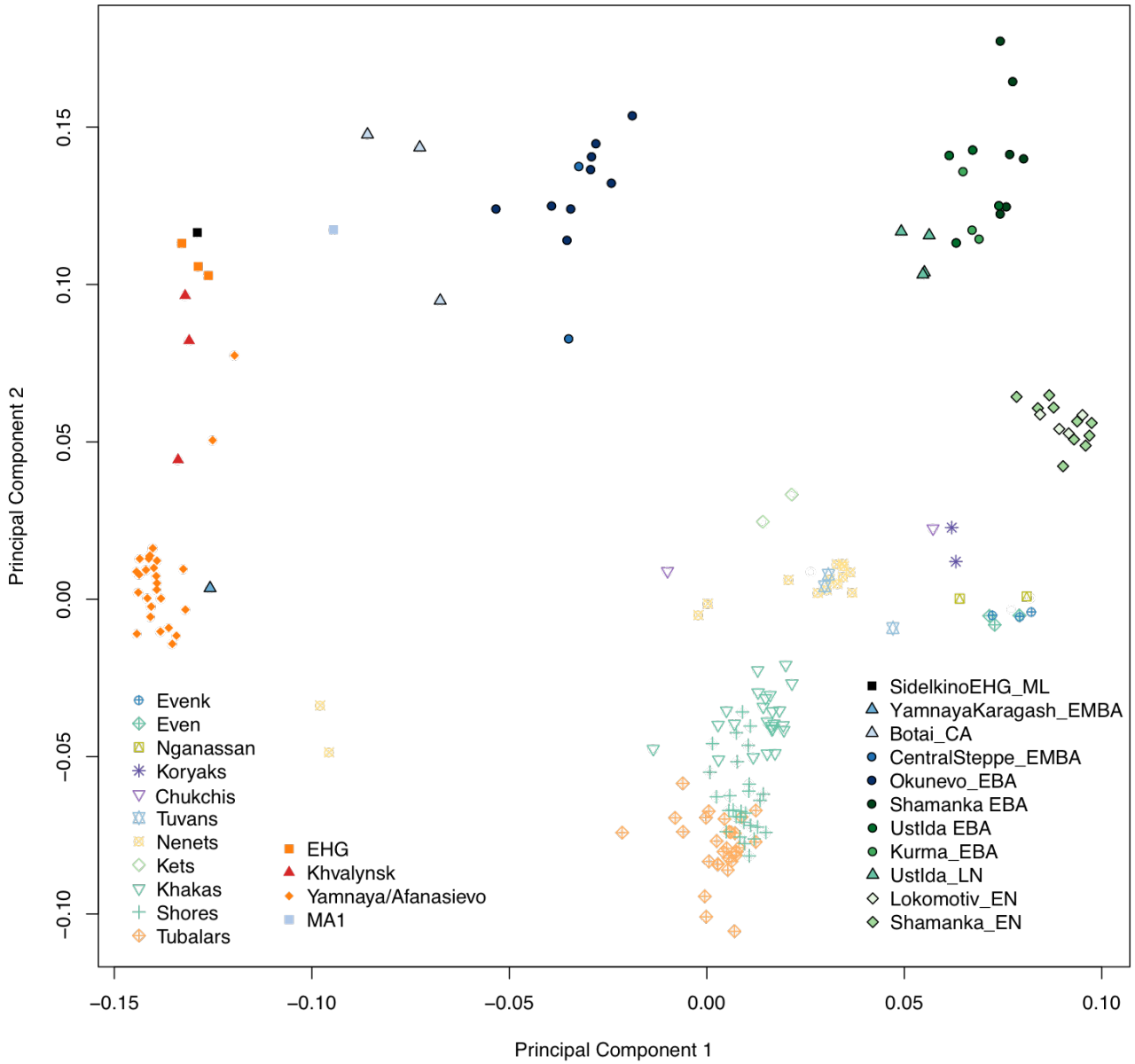

Fig. S23.

PCA estimated with ancient samples from the Steppe and Siberia, together with present-day Siberian populations.

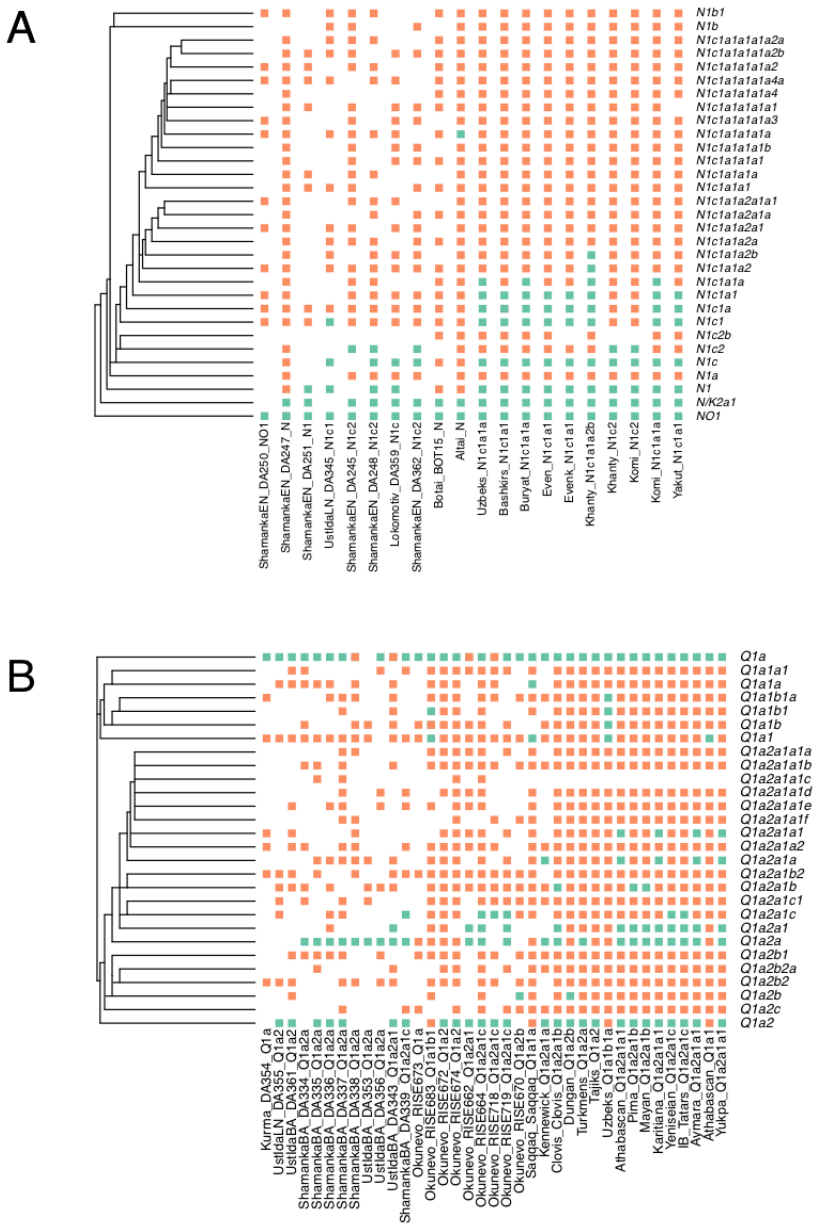

Fig. S24.

Ancient and present-day samples allele status at relevant tips and nodes of the ISOGG Y-chromosomal tree for A) haplogroup N and B) haplogroup Q. Ancestral and derived alleles are represented in orange and green, respectively, and missing data is represented in white. We added tips at relevant nodes of the tree for allowing visualization of ancestral and derived alleles at these.

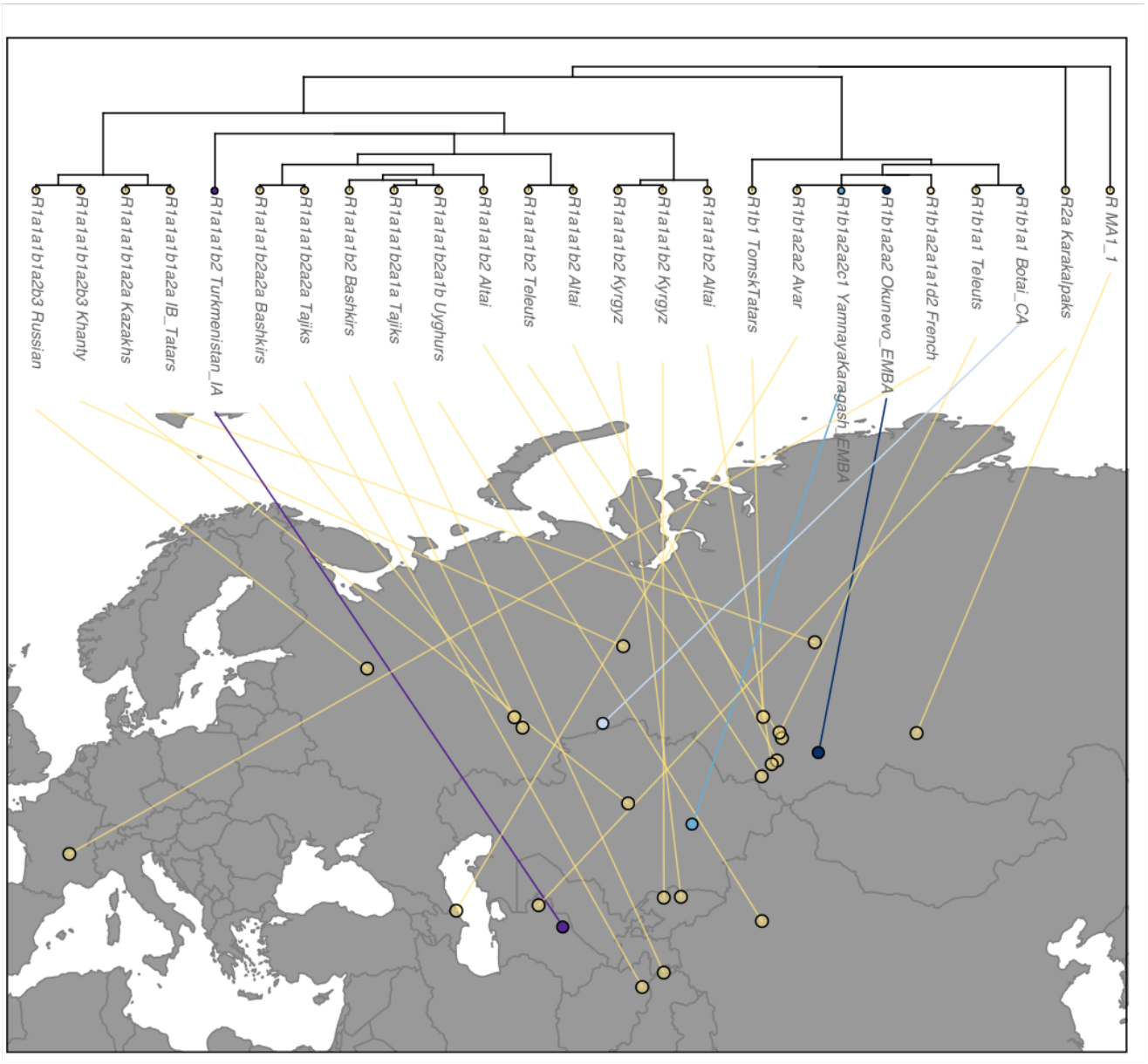

Fig. S25.

Geographical location of ancient samples belonging to major clade R of the Y-chromosome.

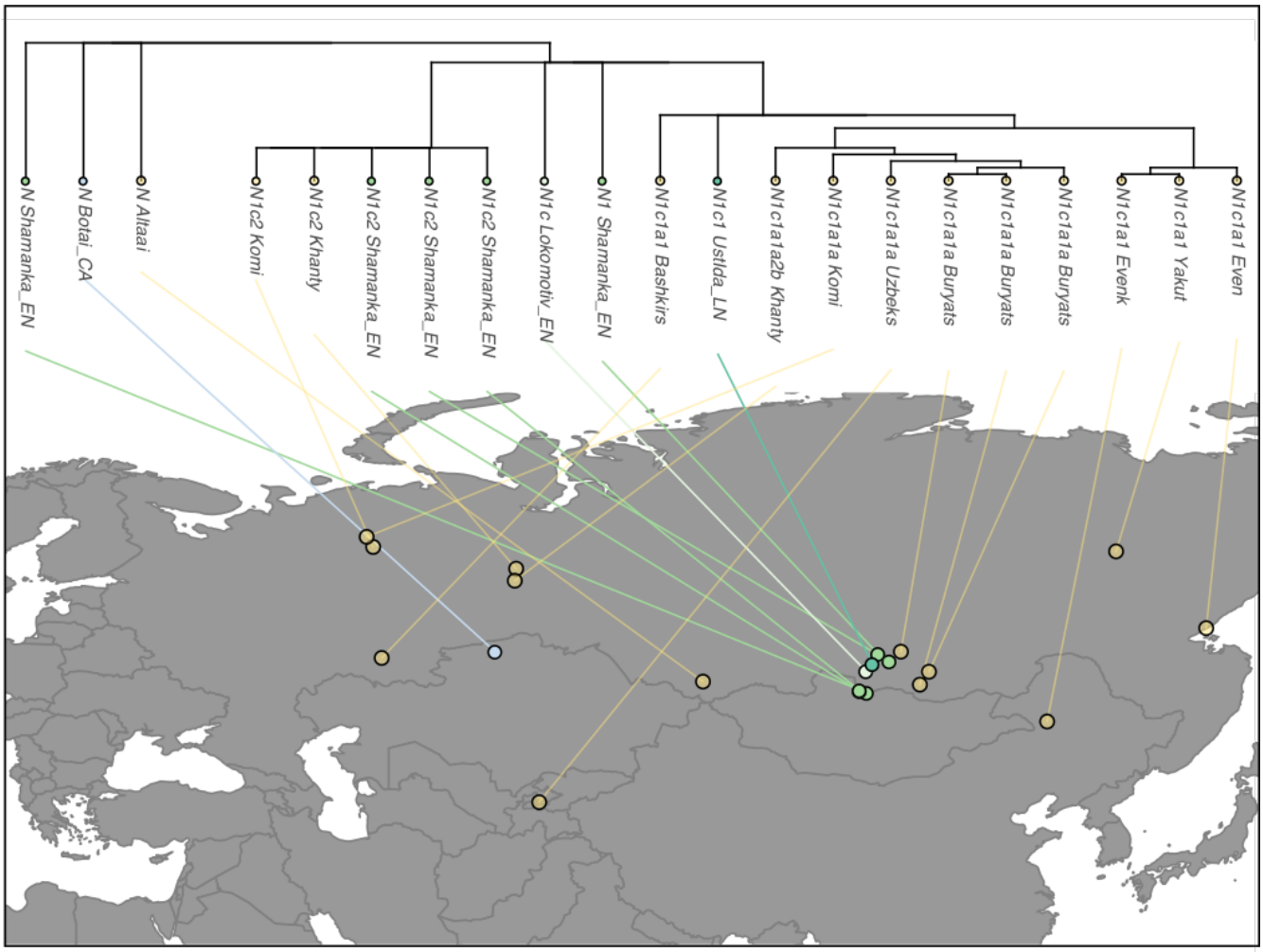

Fig. S26.

Geographical location of ancient samples belonging to major clade N of the Y-chromosome.

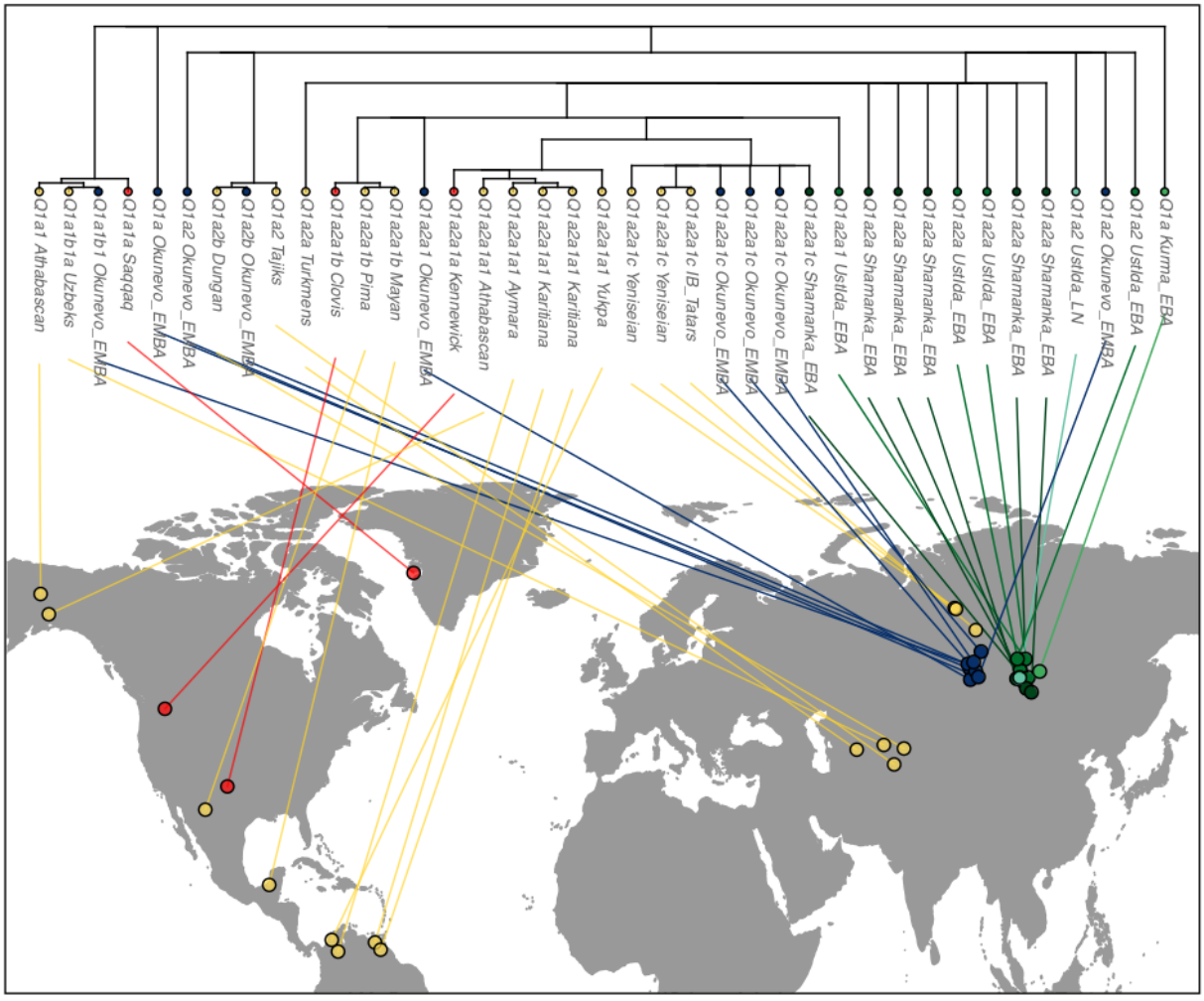

Fig. S27.
Geographical location of ancient samples belonging to major clade Q of the Y-chromosome.

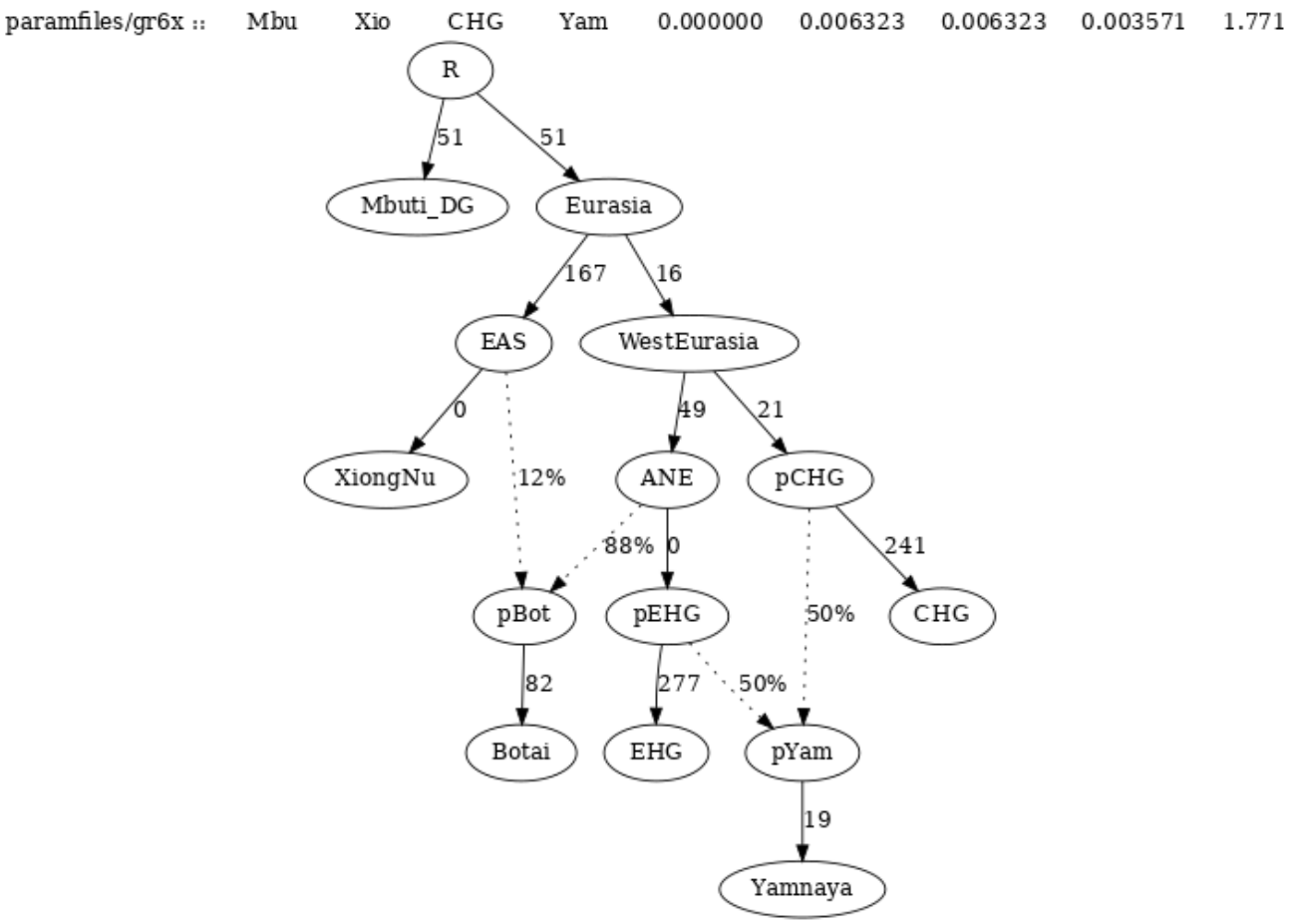

Figure S28.

qpGraph model relating Botai, Yamnaya, and 4 other populations. The model includes no direct Botai-Yamnaya gene flow, and all f4 statistics fit well (|Z| <= 1.77).

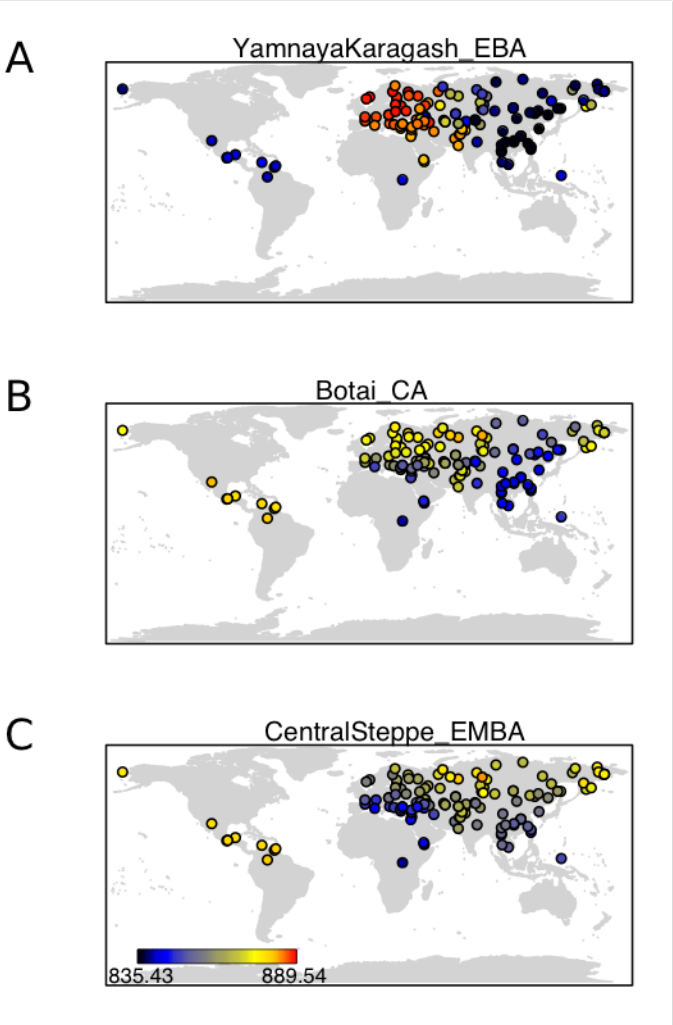

Fig. S29.

Mean haplotype sharing with present-day populations and A) YamnayaKaragash_EBA, B) Botai_CA, and C) Sholpan (CentralSteppe_EMBA).

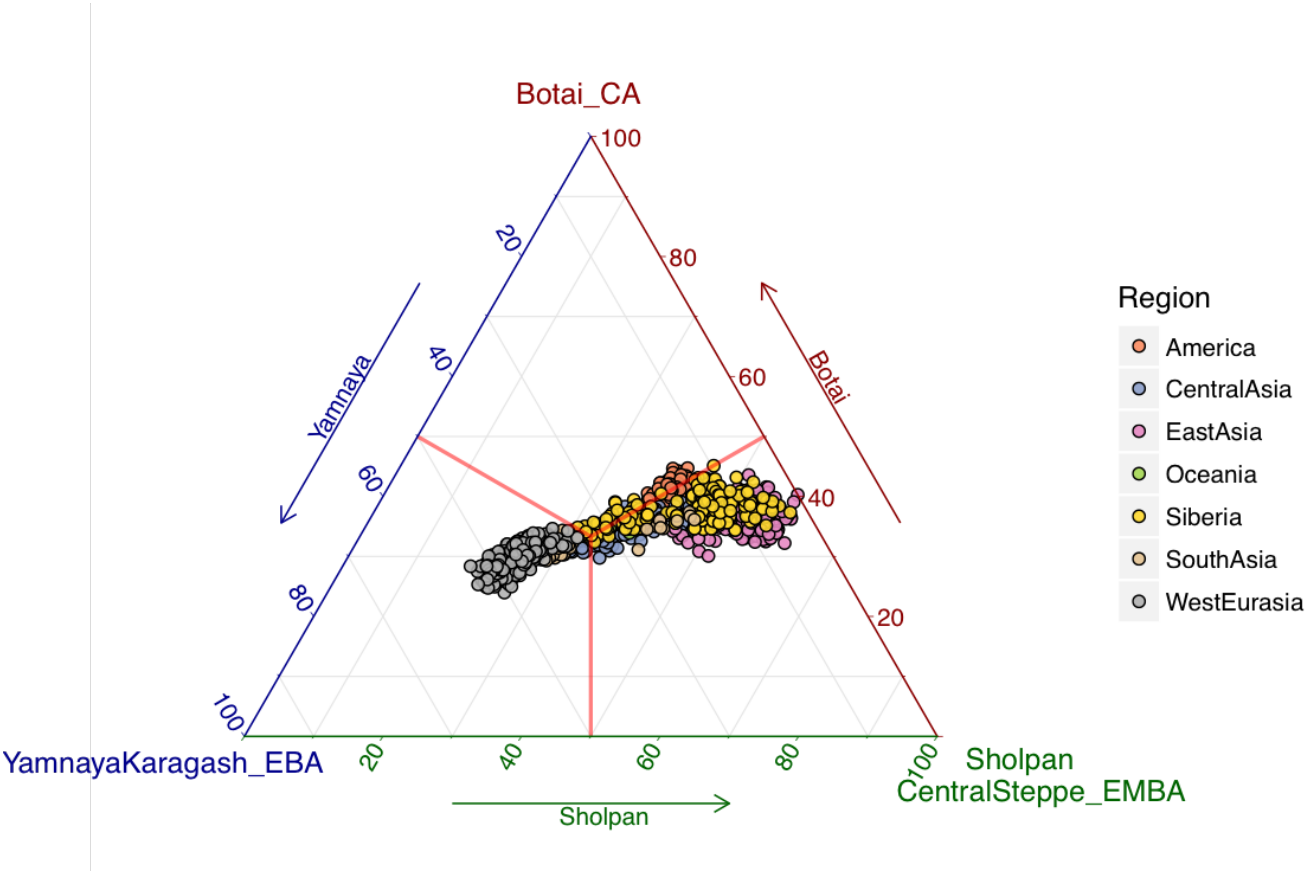

Fig. S30.

Ternary plot of mean haplotype sharing between the high-coverage samples YamnayaKaragash_EBA, Botai_CA, and Sholpan (CentralSteppe_EMBA) with present-day populations.

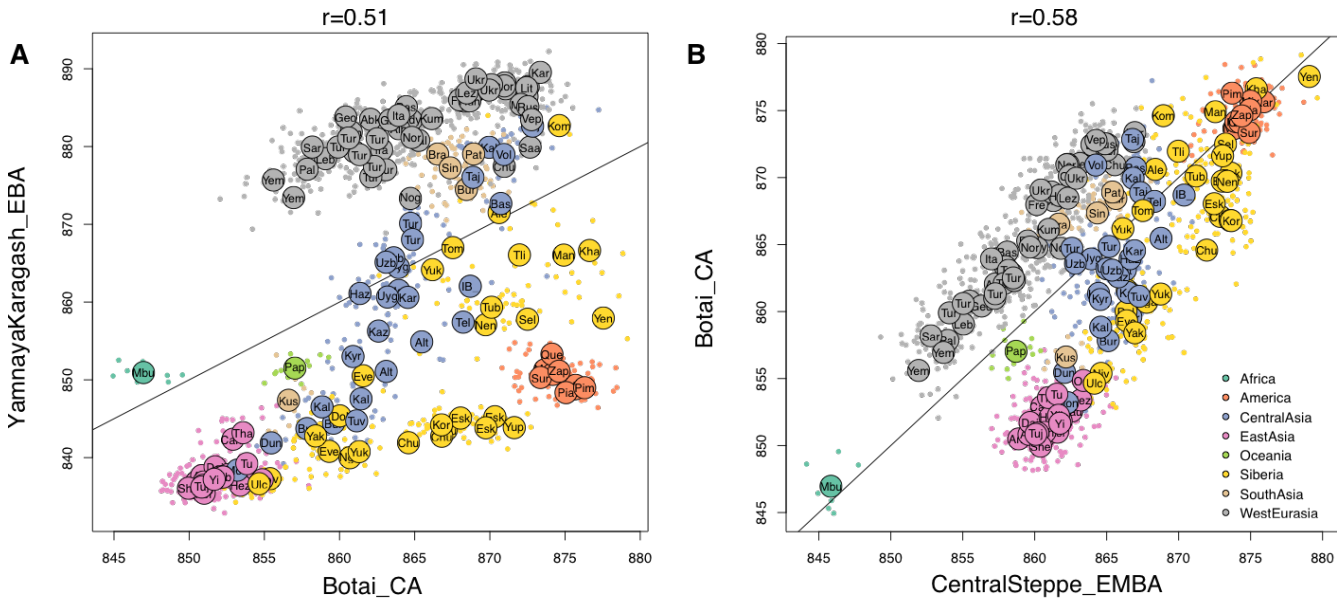

Fig. S31.

Pairwise comparisons of ancient samples in terms of haplotype sharing with present-day populations. A) YamnayaKaragash_EBA and Botai_CA, B) Botai_CA, and Sholpan (CentralSteppe_EMBA).

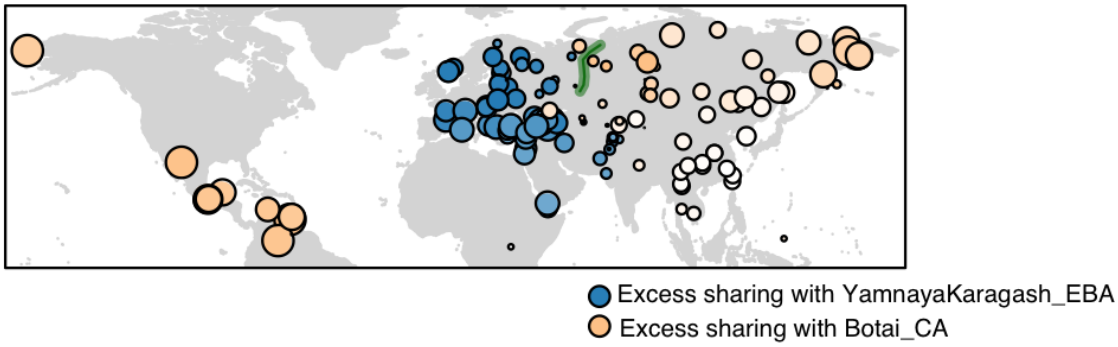

Fig. S32.

Total variation distance comparing Yamnaya and Botai in terms haplotype sharing with modern populations. The color of the circles indicates raw haplotype donation and the size of each circle represents the magnitude of the difference in haplotype sharing between Yamnaya (blue) and Botai (orange). The green line represents the Ural Mountains.

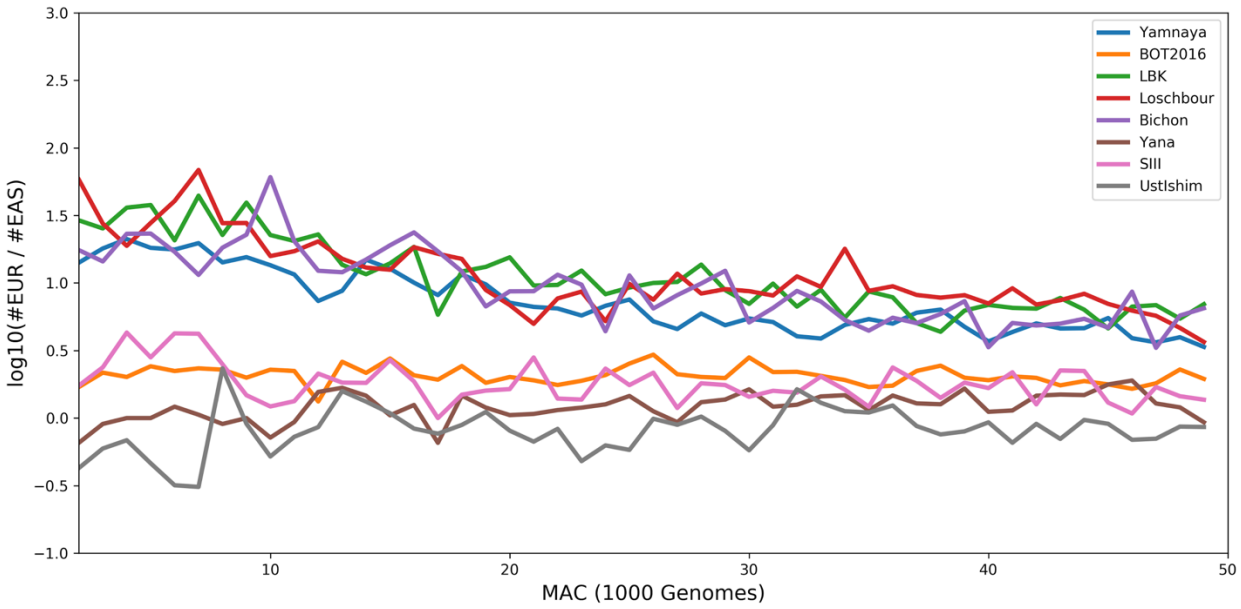

Fig. S33.

Relative numbers of variants that are shared by European (EUR) and East Asian (EAS) populations in the 1000 Genomes Project as a function of minor allele count (MAC) per ancient genome (designated by line color, see legend). Values less than zero indicate higher sharing with East Asian populations (e.g. as seen for Ust-Ishim, gray), and values greater than 0 indicate higher sharing with Europeans (e.g. as seen for Loschbour, red).

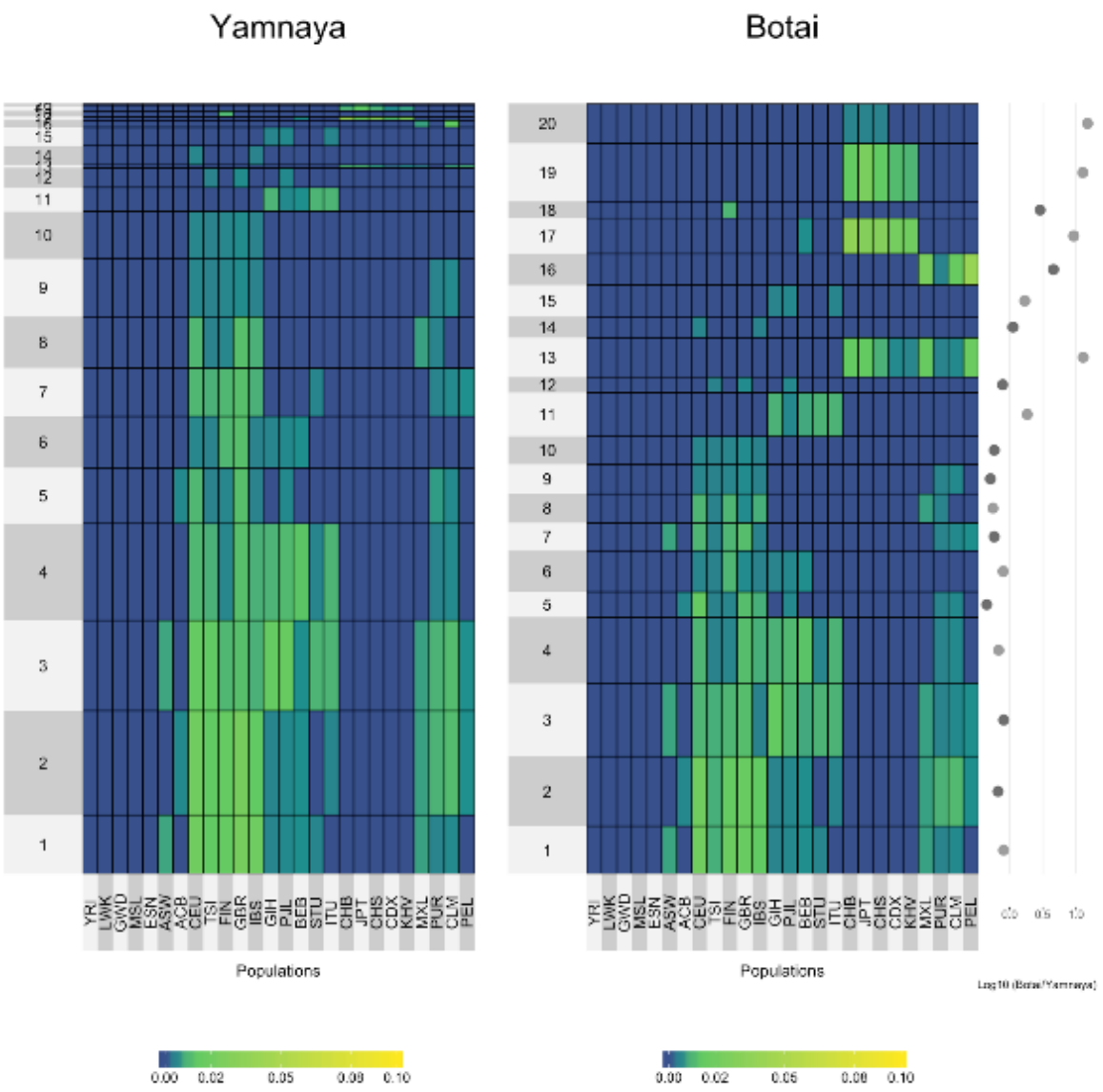

Fig. S34.

The geographic distribution of variants that are shared between modern populations and Yamnaya (left) or Botai (right). Variants have been categorized into 20 discrete geographic patterns. Color intensity represents minor allele frequency, and the relative abundance of each category is represented by breadth along the y-axis. The rightmost panel illustrates the difference in abundance of each category by displaying the (log10) ratio of the fraction of SNPs that fall into that category in Botai vs. Yamnaya. Botai has many more variants that are found in East Asia or East Asia and the Americas (Categories 12, 17, 19, 20). Yamnaya sharing is enriched for variants that are found in Europe-alone or Europe and the Americas (Categories 5, 9). Population labels follow the 1000 Genomes abbreviations.

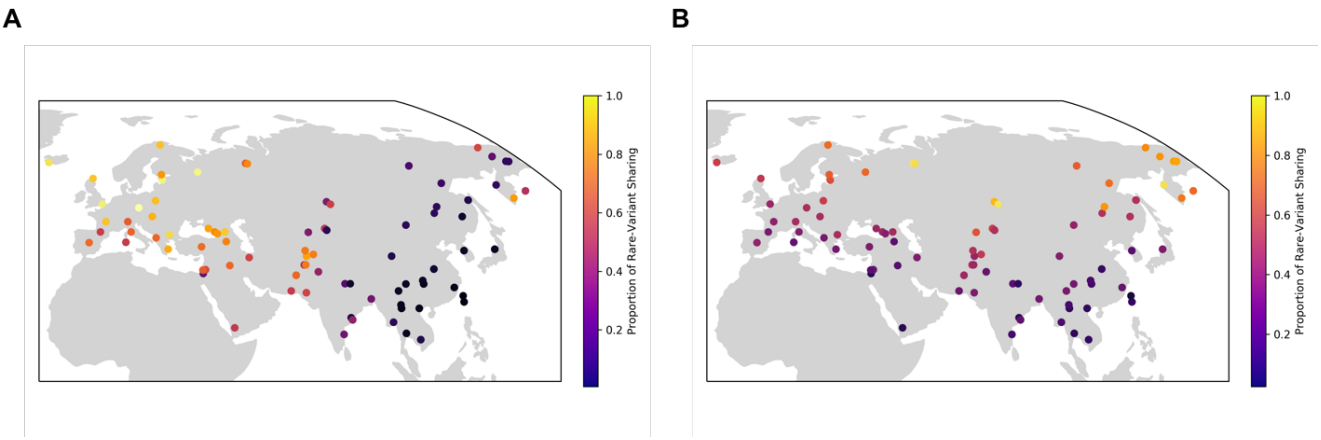

Fig. S35.
Gradients in rare variant sharing between (A) Yamnaya and (B) Botai.

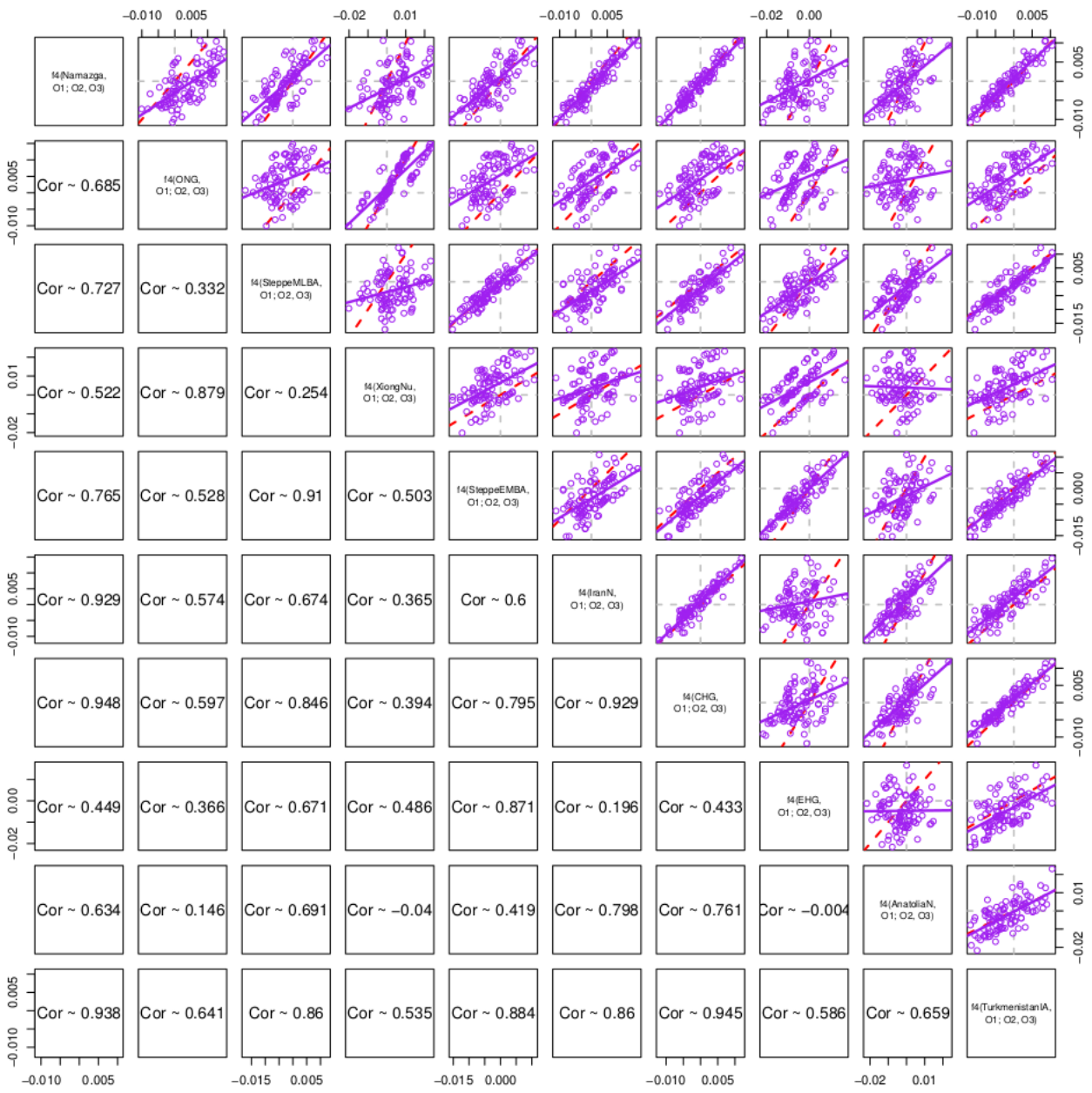

Fig. S36.

Assessment of the information provided by the set of seven outgroups used in the qpAdm models. We computed all possible f_4-statistics of the form *f4(Source$_i$, Outgroup$_1$; Outgroup$_2$, Outgroup$_3$)*, including all the potential sources used in the qpAdm models as well as all possible triplets in the following set of seven outgroups: Mbuti, Ust'Ishim, Clovis, Kostenki14, Switzerland_HG, Natufian, and MA1. For each pair of sources, we plot the corresponding f_4 values in the upper section of the matrix and show the Spearman correlation coefficient in the lower section. Ancestry from sources with high correlation scores will be more difficult to differentiate in qpAdm. We confirm these results using a formal qpWave test (Section S2.10).

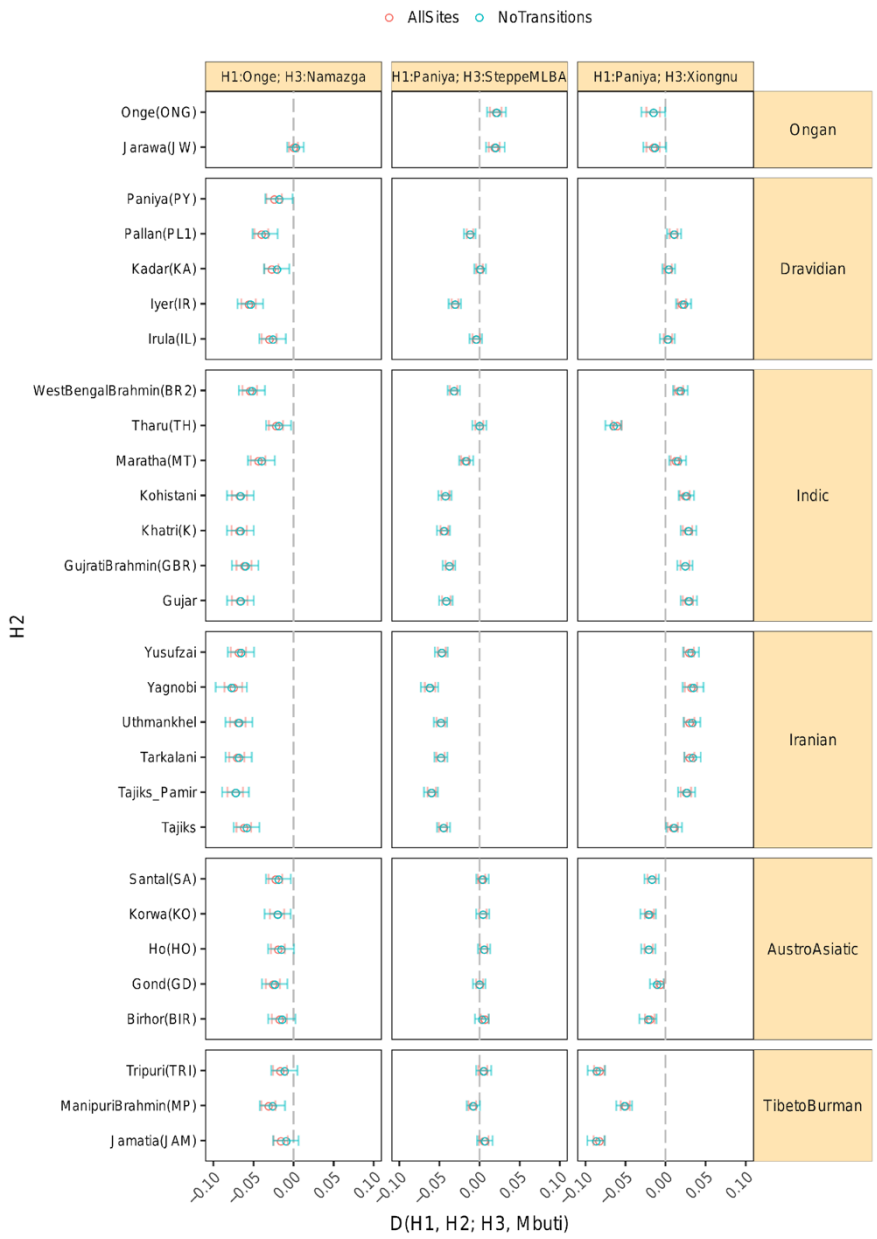

Fig. S37.

D-statistics showing that South Asian populations are consistent with ancestry from 4 sources represented by Onge, Namazga, Late Bronze Age steppe, and Xiongnu nomads (representing East Asians). South Asian populations were grouped according to their language family. For each test, 2 results are shown: one where all sites in the dataset were considered (red points) and one where transition polymorphisms were excluded from the analysis (green points). Positive *D*-statistics indicate that H_1 shares more alleles with H_3 than H_2, while negative statistics indicate that H_2 shares more alleles with H_3 than H_1. Error bars represent ~3.3 standard errors, which corresponds to a *p*-value ~ 0.001.

Sample	Population	Approach	Total reads	Trimmed	Mapped	Endogenous%	Non-duplicate mapped	Clonality %	Coverage	Probability Authentic (95 % CI)
BOT14	Botai	Shotgun/Illumina 2500/HiSeqX10	2192272501	875017599	338021125	40.2	168384462	31.8	3.7	0.9791 (0.9675-0.9864)
BOT15	Botai	Shotgun/Illumina 2500/HiSeqX10	3258239520	1433625118	212931720	14.6	126664850	29	3	0.9980 (0.9863-0.9998)
BOT2016	Botai	Shotgun/Illumina 2500/HiSeqX10	6563324840	2773092327	849998291	30.6	519384291	24	13.6	0.9948 (0.9876-0.9984)
Yamnaya	Yamnaya	Shotgun/Illumina 2500/HiSeqX10	6820561800	3407416890	2675370010	77.3	1123187429	46.1	25.2	0.9886 (0.9827-0.9929)
EBA1	CentralSteppe_EMBA	Shotgun/Illumina 2500/HiSeqX10	3368109128	1434495369	460033579	32.2	257270049	20.6	4.5	0.9566 (0.9439-0.9673)
EBA2	CentralSteppe_EMBA	Shotgun/Illumina 2500/HiSeqX10	2482629018	1191642239	644534121	49.9	394357733	23.5	9.1	0.9937 (0.9873-0.9982)
Sidelkino	SidelkinoEHG_ML	Shotgun/Illumina 2500	374758967	339356731	240884232	71	174337943	4	2.9	0.9954 (0.9885-0.9990)
DA379	Namazga_CA	Shotgun/Illumina 2500	167100464	121785756	3345648	2	3292668	1.6	0.1	0.9855 (0.9288-0.9982)
DA380	Namazga_CA	Shotgun/Illumina 2500	146823356	123045346	29603426	20.2	28144957	4.9	0.5	0.9892 (0.9726-0.9977)
DA381	Namazga_CA	Shotgun/Illumina 2500	301840070	235646784	56568028	18.7	55097492	2.6	0.8	0.9996 (0.9925-0.9999)
DA383	Namazga_CA	Shotgun/Illumina 2500	150392401	128810489	52626463	35	49579801	5.8	0.8	0.9988 (0.9825-0.9997)
DA382	Turkmenistan_IA	Shotgun/Illumina 2500	302180862	277799882	148879656	49.3	145713533	2.1	2.5	0.9996 (0.9940-0.9999)
MA2195	Anatolia_Ottoman	Shotgun/Illumina 2500	139394266	120773729	52093247	37.4	44452475	14.7	0.9	0.9995 (0.9906-0.9999)
MA2196	Anatolia_Ottoman	Shotgun/Illumina 2500	19263518	18517734	12791403	66.4	12328624	3.6	0.3	0.9898 (0.9692-0.9983)
MA2197	Anatolia_IA	Shotgun/Illumina 2500	14202511	11907084	6703251	47.2	6367699	5	0.1	0.9983 (0.9671-0.9997)
MA2198	Anatolia_IA	Shotgun/Illumina 2500	68947699	66868570	38282336	55.5	36731227	4.1	0.8	0.9832 (0.9703-0.9909)
MA2200	Anatolia_MLBA	Shotgun/Illumina 2500	278699922	234951794	135920736	48.8	121592180	10.5	2.2	0.9877 (0.9771-0.9944)
MA2203	Anatolia_MLBA	Shotgun/Illumina 2500	200867597	169526141	52589100	26.2	49614004	5.7	0.9	0.9996 (0.9928-0.9999)
MA2205	Anatolia_MLBA	Shotgun/Illumina 2500	85020838	79162433	44015910	51.8	42080032	4.4	0.8	0.9995 (0.9924-0.9999)
MA2206	Anatolia_MLBA	Shotgun/Illumina 2500	248367711	214763480	40763269	16.4	22308224	45.3	0.4	0.9943 (0.9843-0.9992)
MA2208	Anatolia_MLBA	Shotgun/Illumina 2500	125464977	67512900	27649991	22	6810359	75.4	0.1	0.9449 (0.8869-0.9819)
MA2210	Anatolia_EBA	Shotgun/Illumina 2500	126406011	109308921	50150042	39.7	47737328	4.8	0.9	0.9997 (0.9939-0.9999)
MA2212	Anatolia_EBA	Shotgun/Illumina 2500	239127212	149093443	64936710	27.2	59049012	9.1	0.9	0.9989 (0.9895-0.9998)
MA2213	Anatolia_EBA	Shotgun/Illumina 2500	368438736	277789904	72819648	19.8	68976171	5.3	1.2	0.9944 (0.9820-0.9995)
RISE515	Okunevo_EMBA	Shotgun/Illumina 2500	1783268077	1461556032	199713542	13.7	26762962	86.6	0.6	0.9998 (0.9954-1.0000)
RISE516	Okunevo_EMBA	Shotgun/Illumina 2500	1278867523	1014478964	138968978	13.7	44121941	68.3	0.9	0.9998 (0.9957-1.0000)
RISE662	Okunevo_EMBA	Shotgun/Illumina 2500	225121907	166283480	102779809	61.8	46334917	54.9	0.6	0.9997 (0.9932-0.9999)
RISE664	Okunevo_EMBA	Shotgun/Illumina 2500	389290051	364213753	254781563	70	233010348	8.5	4.6	0.9525 (0.9348-0.9650)
RISE667	Okunevo_EMBA	Shotgun/Illumina 2500	54173973	49772083	7318294	14.7	10279306	31	0.2	0.9994 (0.9879-0.9999)
RISE670	Okunevo_EMBA	Shotgun/Illumina 2500	108012364	100845664	32658656	32.4	31326486	4.1	0.7	0.9984 (0.9824-0.9998)
RISE671	Okunevo_EMBA	Shotgun/Illumina 2500	550866525	476213513	15353684	3.2	14693369	4.3	0.3	0.9997 (0.9933-1.0000)
RISE672	Okunevo_EMBA	Shotgun/Illumina 2500	414291899	312321446	95421239	30.6	75136424	21.3	1.2	0.9992 (0.9907-0.9999)
RISE673	Okunevo_EMBA	Shotgun/Illumina 2500	144799175	132936270	5719317	4.3	5701246	41.4	0.1	0.9951 (0.9776-0.9994)
RISE674	Okunevo_EMBA	Shotgun/Illumina 2500	245701496	239518662	136219559	56.9	125799394	7.6	2.6	0.9992 (0.9916-0.9999)
RISE675	Okunevo_EMBA	Shotgun/Illumina 2500	64309347	61733688	31770813	51.5	23328642	26.6	0.5	0.9992 (0.9845-0.9999)
RISE677	Okunevo_EMBA	Shotgun/Illumina 2500	113279212	109527506	3671588	3.4	12617825	34.9	0.3	0.9993 (0.9882-0.9999)
RISE680	Okunevo_EMBA	Shotgun/Illumina 2500	570996627	495906372	80009784	16.1	72927202	8.9	1.5	0.9997 (0.9946-1.0000)
RISE681	Okunevo_EMBA	Shotgun/Illumina 2500	155122917	137120090	23342542	17	24992043	80.4	0.5	0.9979 (0.9925-0.9997)
RISE683	Okunevo_EMBA	Shotgun/Illumina 2500	459719628	420130869	154765683	36.8	123037686	20.5	2	0.9995 (0.9909-0.9999)
RISE684	Okunevo_EMBA	Shotgun/Illumina 2500	57184624	53344713	25467314	47.7	24747517	2.8	0.5	0.9841 (0.9628-0.9953)
RISE685	Okunevo_EMBA	Shotgun/Illumina 2500	267430858	231916916	100594496	43.4	62567628	37.8	1.3	0.9739 (0.9586-0.9837)
RISE718	Okunevo_EMBA	Shotgun/Illumina 2500	78903540	75346364	46257278	61.4	42442137	8.2	0.8	0.9780 (0.9611-0.9897)
RISE719	Okunevo_EMBA	Shotgun/Illumina 2500	68022374	66251023	30840649	46.6	26628029	13.6	0.6	0.9874 (0.9760-0.9940)
DA245	Shamanka_EN	Shotgun/Illumina 2500	162814301	157893611	107620040	68.2	101495484	5.7	2.2	0.9945 (0.9861-0.9992)
DA246	Shamanka_EN	Shotgun/Illumina 2500	196808425	192033838	142597590	74.3	129711403	9	2.9	0.9942 (0.9840-0.9992)
DA247	Shamanka_EN	Shotgun/Illumina 2500	227346763	218167045	123110325	56.4	113828443	7.5	2.4	0.9772 (0.9644-0.9855)
DA248	Shamanka_EN	Shotgun/Illumina 2500	176878767	169341415	116361008	68.7	104983472	9.8	2.3	0.9763 (0.9613-0.9876)
DA249	Shamanka_EN	Shotgun/Illumina 2500	327775553	318997226	230221567	72.2	204074308	11.4	4.5	0.9642 (0.9521-0.9740)
DA250	Shamanka_EN	Shotgun/Illumina 2500	154098108	140930953	57434613	40.8	54052423	5.9	0.9	0.9962 (0.9826-0.9997)
DA251	Shamanka_EN	Shotgun/Illumina 2500	232916764	210122856	26853705	12.8	25878772	3.6	0.6	0.9879 (0.9739-0.9976)
DA252	Shamanka_EN	Shotgun/Illumina 2500	200437910	192696563	117408213	60.9	108012064	8	2.4	0.9839 (0.9663-0.9949)
DA253	Shamanka_EN	Shotgun/Illumina 2500	218275178	209560446	139008225	66.3	126504028	9	2.7	0.9969 (0.9916-0.9993)
DA334	Shamanka_EBA	Shotgun/Illumina 2500	103757960	95727037	26361756	27.5	24589270	6.7	0.5	0.9683 (0.9536-0.9808)
DA335	Shamanka_EBA	Shotgun/Illumina 2500	81631491	76980076	34539800	44.9	22685723	34.3	0.5	0.9987 (0.9936-0.9997)
DA336	Shamanka_EBA	Shotgun/Illumina 2500	97815430	91827389	40179131	43.8	32714889	18.6	0.7	0.9758 (0.9620-0.9848)
DA337	Shamanka_EBA	Shotgun/Illumina 2500	105263805	98281097	62573101	63.7	56238516	10.1	1.1	0.9816 (0.9714-0.9891)
DA338	Shamanka_EBA	Shotgun/Illumina 2500	94862480	88118581	23377137	26.5	20699277	11.5	0.4	0.9774 (0.9629-0.9892)
DA339	Shamanka_EBA	Shotgun/Illumina 2500	72483521	69795739	15572767	22.3	14097198	9.5	0.3	0.9929 (0.9819-0.9993)
DA340	Lokomotiv_EN	Shotgun/Illumina 2500	208345480	188293011	31481459	16.7	30683122	2.5	0.6	0.9993 (0.9863-0.9999)
DA341	Lokomotiv_EN	Shotgun/Illumina 2500	118127945	110691995	59780536	54	56135151	6.1	1.2	0.9945 (0.9826-0.9993)
DA342	UstIda_LN	Shotgun/Illumina 2500	195311656	186638454	90564181	48.5	83344165	8	1.7	0.9947 (0.9846-0.9991)
DA343	UstIda_EBA	Shotgun/Illumina 2500	76731250	71099845	34742212	48.9	29838435	14.1	0.6	0.9914 (0.9794-0.9989)
DA344	UstIda_LN	Shotgun/Illumina 2500	50563357	46552213	24151745	51.9	11222522	53.5	0.2	0.9826 (0.9599-0.9941)
DA345	UstIda_LN	Shotgun/Illumina 2500	93807114	88926771	53959672	60.7	49305041	8.6	1	0.9543 (0.9344-0.9693)
DA353	UstIda_EBA	Shotgun/Illumina 2500	72655137	60926879	38003425	62.4	10048972	73.6	0.2	0.9502 (0.8794-0.9834)
DA354	Kurma_EBA	Shotgun/Illumina 2500	167216578	126180424	17777366	14.1	12815167	27.9	0.2	0.9987 (0.9729-0.9998)
DA355	UstIda_LN	Shotgun/Illumina 2500	150852349	125317315	36542206	29.2	27240982	25.5	0.4	0.9990 (0.9828-0.9998)
DA356	UstIda_EBA	Shotgun/Illumina 2500	146021057	132174503	31814087	24.1	20058229	37	0.4	0.9612 (0.9434-0.9732)
DA357	Lokomotiv_EN	Shotgun/Illumina 2500	116412845	106799820	32852067	30.8	19213306	41.5	0.4	0.9992 (0.9870-0.9998)
DA358	Kurma_EBA	Shotgun/Illumina 2500	197490686	163807739	55746327	34	44528471	20.1	0.9	0.9996 (0.9924-0.9999)
DA359	Lokomotiv_EN	Shotgun/Illumina 2500	94856162	86833551	50797883	58.5	35078164	30.9	0.7	0.9989 (0.9851-0.9998)
DA360	Kurma_EBA	Shotgun/Illumina 2500	110520252	94120580	27668547	29.4	15883302	42.6	0.3	0.9985 (0.9848-0.9998)
DA361	UstIda_EBA	Shotgun/Illumina 2500	168686829	141465106	49697236	35.1	28622309	42.4	0.6	0.9556 (0.9366-0.9680)
DA362	Shamanka_EN	Shotgun/Illumina 2500	156194048	148324489	99610247	67.2	50570847	49.2	1.1	0.9953 (0.9872-0.9994)

Table S1.

Details of the sequence data generated in the present study including read number before and after filtering, extent of duplication, genomic coverage, sex, and contamination estimates.

Cemetery	*EN*	*LN*	*EBA*	*Total*
Lokomotiv	4			4
Shamanka II	10		6	16
Ust'Ida		4	4	8
Kurma			3	3
Total	14	4	13	31

Table S2.

Summary of human tooth samples submitted for the analysis reported in the current paper.

Newly sequenced samples					
Population/ Sample Label	Geographical Range	Period	Approximate time before present	Subsistence	Sample size
SidelkinoEHG_ML	Eastern Europe	Mesolithic	11500–11000	Hunter-Gatherer	1
Botai_CA	Central Steppe	Copper Age	5500–5300	Hunter-Herder	3
YamnayaKaragash_EBA	Central Steppe	Early Bronze Age	4900	Pastoral	1
CentralSteppe_EMBA	Central Steppe	Early/Middle Bronze Age	4200	Unknown/mixed	2
Okunevo_EMBA	Minusinsk Basin	Early/Middle Bronze Age	4500–4000	Mixed HG / Pastoralist	19
Shamanka_EN	Cis-Baikal	Early Neolithic	7200–6200	Hunter-Gatherer	12
Lokomotiv_EN	Cis-Baikal	Early Neolithic	6700	Hunter-Gatherer	4
UstIda_LN	Cis-Baikal	Late Neolithic	5000	Hunter-Gatherer	4
Kurma_EBA	Cis-Baikal	Early Bronze Age	4200–4000	Hunter-Gatherer	3
Shamanka_EBA	Cis-Baikal	Early Bronze Age	4000–3800	Hunter-Gatherer	4
UstIda_EBA	Cis-Baikal	Early Bronze Age	4000–3800	Hunter-Gatherer	4
Namazga_CA	Turkmenistan	Copper Age	5300–5200	Agriculture	4
Turkmenistan_IA	Turkmenistan	Iron Age	2800	Agriculture	1
Anatolia_EBA	Central Anatolia	Early Bronze Age	4200	Agriculture	3
Anatolia_MLBA	Central Anatolia	Middle/Late Bronze Age	3600	Agriculture	5
Anatolia_IA	Central Anatolia	Iron Age	2600	Agriculture	2
Anatolia_Ottoman	Central Anatolia	Late Medieval	500	Agriculture	2

Previously published samples mentioned throughout the manuscript					
Population/Sample Label	Geographical Range	Period	Approximate time before present	Subsistence	Reference
MA1	Cis-Baikal	Paleolithic	24423–23891	Hunter-Gatherer	(*26*)
AG-2	south central Siberia	Paleolithic	17075–16750	Hunter-Gatherer	(*26*)
AG-3	south central Siberia	Paleolithic	14710	Hunter-Gatherer	(*27*)
CHG	Caucasus	Upper Paleolithic-Mesolithic	13300–9700	Hunter-Gatherer	(*7*)
Natufian	Levant	Epipaleolithic	13840–11760	Hunter-Gatherer	(*42*)
EHG	Eastern Europe	Mesolithic	8850–7000	Hunter-Gatherer	(*2, 47*)
Iran_N	Iran	Neolithic	9950–9650	Hunter-Gatherer	(*42*)
Steppe_EMBA	Eastern Europe/Altai	Early /Middle Bronze Age	5000–4500*	Nomadic/Pastoral	(*1, 2*)
Steppe_MLBA	Eastern Europe & Central Asia	Middle/Late Bronze Age	4500–3200	Nomadic/Pastoral	(*1, 2*)
Xiongu_IA	Eastern Steppe	Iron Age	2300–1900	Nomadic/Pastoral	(*3*)
Steppe_Eneolithic	Western Steppe	Copper Age	7150–5950	Nomadic/Pastoral	(*47*)
Armenian Copper Age	Armenia	Copper Age	5397–5230	Agriculture	(*42*)
Devil's Gate	East Asia	Early Neolithic	7700	Hunter-Gatherer	(*25*)
Iran_Chl	Iran	Copper Age	5900–5040	Agriculture	(*42*)
Anatolia_N	Anatolia	Neolithic	8350–7550	Agriculture	(*42*)
Anatolia_ChL	Anatolia	Copper Age	5900–5700	Agriculture	(*42*)
Iran LN	Iran	Late Neolithic	6850	Agriculture	(*42*)
Ust-Ishim	Siberia	Paleolithic	45000	Hunter-Gatherer	(*128*)
BR2	Hungary	Bronze Age	3220–3060	Agriculture?/Mixed?	(*139*)
Clovis	Americas	Paleolithic	13000–12600	Hunter-Gatherer	(*137*)
Kennewick	Americas	Paleolithic	8340–9200	Hunter-Gatherer	(*142*)
Saqqaq	Americas	Arctic Small Tool	4170–3600	Hunter-Gatherer	(*138*)

* including two genetic outliers from the Middle Bronze Age

Table S4. Overview of population labels and population sizes of groups newly sequenced and of relevant published samples referred to throughout the manuscript.

Individual 1	Individual 2	k0_hat	k1_hat	k2_hat	pi_HAT	Number of SNPs
DA336	DA338	0.165	0.491	0.344	0.589	57548
RISE515	RISE673	0.362	0.319	0.319	0.478	18932
DA379	DA380	0.225	0.634	0.141	0.458	6711
RISE516	RISE672	0.365	0.451	0.185	0.41	129263
DA334	DA335	0.228	0.767	0.005	0.388	51785
RISE671	RISE673	0.537	0.32	0.143	0.303	11542
DA340	DA341	0.488	0.443	0.069	0.29	92418
RISE515	RISE671	0.738	0.01	0.252	0.257	40957
RISE670	RISE674	0.521	0.454	0.025	0.252	141703
RISE662	RISE664	0.539	0.432	0.029	0.245	134407
DA353	DA361	0.663	0.194	0.143	0.24	23917
RISE515	RISE516	0.573	0.392	0.035	0.231	92488
RISE515	RISE667	0.602	0.332	0.065	0.231	30717
RISE515	RISE672	0.608	0.335	0.057	0.224	96837
RISE515	RISE674	0.594	0.371	0.035	0.221	125410
RISE672	RISE673	0.751	0.077	0.172	0.211	26883
RISE667	RISE673	0.654	0.285	0.062	0.204	8577

Table S5.

Highest values obtained in the analysis of pairwise relatedness with LCMLKIN.

Param	Inferred	95% CI	Bias	SD
$P_{\text{yamFromCHG}}$	0.54	(0.30, 0.73)	-0.03	0.11
T_{admixYam}	4900.01	(4900.01, 19057.78)	2744.71	4140.22
$T_{\text{KK1-YamCHG}}$	26815.73	(9825.75, 30891.50)	-3273.67	5401.31
$T_{\text{Sid-YamANE}}$	11240.14	(11240.02, 19058.17)	1166.89	2053.84
$T_{\text{Botai-YamANE}}$	17140.53	(11238.49, 22830.86)	-951.31	3071.72
$T_{\text{Botai-KK1}}$	38133.78	(33804.71, 41742.71)	-674.86	2056.11
N_{Botai}	3666.05	(1420.77, 5447.87)	-363.32	1090.17
N_{KK1}	1765.31	(17.18, 2178.15)	-289.48	535.25
$N_{\text{Ancestral}}$	11404.65	(11257.32, 11518.16)	-22.78	64.90
N_{Eurasia}	3846.61	(3573.90, 4570.13)	133.05	264.97

Table S6.

Point estimates for the model in Fig. S16, along with parametric bootstrap estimates of 95% confidence intervals, bias, and standard deviation.

X	Expected	Observed	Z-score
KK1	-0.209	-0.197	0.644
AncestralAllele	-0.308	-0.337	-1.834

Table S7.

f_4^*(Yamnaya, Sidelkino, Botai, X) residuals for the model in Fig. S16. Z-scores were computed using a block-jackknife with 100 blocks.

Param	Inferred	95% CI	Bias	SD
N_{Mbuti}	23290.67	(22464.22, 24119.81)	18.11	414.07
N_{Steppe}	3563.47	(2882.31, 4352.43)	-8.47	363.63
N_{Botai}	2741.21	(1481.46, 3999.45)	-212.05	614.33
N_{Sholpan}	1267.10	(958.19, 1607.68)	4.91	157.74
N_{ANE}	2159.74	(1866.16, 2335.58)	-75.66	113.54
N_{Han}	5793.64	(5530.65, 6026.03)	-24.73	130.35
$N_{\text{Ancestral}}$	12464.47	(12412.77, 12506.72)	-1.38	24.13
N_{Eurasia}	3683.89	(3616.90, 3777.24)	10.45	45.70
$T_{\text{Mbuti-Eurasia}}$	122513.04	(121630.09, 123547.00)	34.46	513.67
$T_{\text{AEA-ANE}}$	48294.75	(46577.94, 49632.08)	-129.92	754.35
$T_{\text{Han-ShamankaEN}}$	17486.66	(16652.49, 18268.40)	-40.27	390.92
$T_{\text{MA1-GhostANE}}$	26580.47	(25319.17, 29126.32)	591.97	896.66
$T_{\text{Steppe-GhostANE}}$	12536.92	(11607.70, 19220.01)	2174.32	1874.60
$T_{\text{AEA->Steppe}}$	12535.30	(10924.24, 13976.98)	-119.99	813.02
$P_{\text{AEA->Steppe}}$	0.51	(0.49, 0.53)	0.00	0.01
$T_{\text{Sholpan-Okunevo}}$	10545.15	(9528.28, 11378.47)	-101.47	490.01
T_{Botai}	11358.50	(8893.71, 12411.86)	-735.59	905.34
$P_{\text{GhostANE->Botai}}$	0.40	(0.35, 0.43)	-0.00	0.02

Table S8.

Point estimates for the model in Fig. S19, along with parametric bootstrap estimates of 95% confidence intervals, bias, and standard deviation.

X	Expected	Observed	Z-score
Mbuti	-0.117	-0.132	-0.924
Okunevo	-0.054	-0.130	-5.585
Sholpan	-0.031	-0.065	-1.830
MA1	-0.063	-0.074	-0.663
Han	-0.124	-0.132	-0.463
ShamankaEN	-0.136	-0.129	0.481
KK1	-0.182	-0.200	-0.855
AncestralAllele	-0.128	-0.152	-1.482

Table S9.

f_4^*(Yamnaya, Sidelkino, Botai, X) residuals for the model in Fig. S20. Z-scores were computed using a block-jackknife with 100 blocks.

Param	Inferred	95% CI	Bias	SD
N_{Mbuti}	22419.38	(21720.85, 23040.01)	2.10	338.59
N_{Steppe}	952.26	(844.45, 1479.24)	134.37	168.25
N_{Botai}	991.05	(507.29, 2045.61)	56.91	364.35
N_{Sholpan}	593.79	(432.98, 886.85)	41.78	112.93
N_{ANE}	2700.77	(2555.46, 2890.90)	14.38	80.78
N_{Han}	4852.23	(4652.50, 5031.59)	11.83	93.47
$N_{\text{Ancestral}}$	12442.81	(12404.26, 12507.90)	10.32	26.28
$T_{\text{Mbuti-Eurasia}}$	121290.98	(120399.04, 122237.33)	32.75	474.59
N_{Eurasia}	4039.99	(3961.31, 4109.05)	-6.35	37.64
$T_{\text{AEA-ANE}}$	42394.72	(41573.03, 43443.26)	159.31	453.13
$T_{\text{Han-ShamankaEN}}$	15250.68	(14696.51, 16021.19)	88.97	321.32
$T_{\text{MA1-GhostANE}}$	24000.02	(24000.02, 24003.13)	6.79	63.62
$T_{\text{Steppe-GhostANE}}$	7752.34	(7000.21, 12526.69)	500.88	1284.48
$T_{\text{AEA->Steppe}}$	7270.62	(7000.01, 8223.71)	64.31	324.65
$P_{\text{AEA->Steppe}}$	0.45	(0.41, 0.45)	-0.02	0.01
$T_{\text{Sholpan-Okunevo}}$	6840.96	(6521.14, 7794.35)	153.18	335.92
T_{Botai}	7262.52	(6757.08, 8126.88)	18.86	331.07
$P_{\text{GhostANE->Botai}}$	0.49	(0.41, 0.52)	-0.03	0.03
$T_{\text{Sidelkino-GhostANE}}$	16424.22	(14904.99, 16997.49)	-440.82	493.71
N_{CHG}	5728.53	(3011.24, 10964.63)	368.41	2110.17
$T_{\text{CHG-ANE}}$	36010.93	(35293.54, 36814.39)	53.88	384.89
$P_{\text{Sidelkino->Yamnaya}}$	0.40	(0.23, 0.57)	-0.00	0.08
N_{KK1}	1053.53	(27.90, 1883.24)	-23.60	490.27

$T_{\text{KK1-YamnayaCHG}}$	20032.10	(9965.85, 28033.62)	-194.90	4803.73
$P_{\text{Yamnaya->Okunevo}}$	0.16	(0.09, 0.14)	-0.05	0.01

Table S10.

Estimated parameters for the final model in Fig. S21, along with parametric bootstrap estimates of 95% confidence intervals, bias, and standard deviation.

X	Expected	Observed	Z-score
Mbuti	-0.148	-0.132	1.060
Okunevo	-0.112	-0.130	-1.349
Sholpan	-0.046	-0.065	-1.009
MA1	-0.059	-0.074	-0.868
Han	-0.156	-0.132	1.399
ShamankaEN	-0.169	-0.129	2.702
KK1	-0.230	-0.200	1.488
AncestralAllele	-0.159	-0.152	0.445

Table S11.

f_4^*(Yamnaya, Sidelkino, Botai, *X*) residuals for the model in Fig. S21. Z-scores were computed using a block-jackknife with 100 blocks.

Target	Source 1	Source 2	Source 3	tandard Error	tandard Error	tandard Error	P-value
	EHG	CHG					
Steppe_EMBA	0.53	0.47		0.04	0.04		0.12
	EHG	CHG	Botai_CA				
Steppe_EMBA	0.55	0.48	-0.03	0.07	0.04	-0.03	0.06
	Baikal_EN	MA1					
Baikal_LNBA	0.92	0.08		0.03	0.03		0.32

Table S12. qpAdm results calculated using 6 outgroups (Mbuti.DG, Ust_Ishim, Clovis, Kostenki14, Switzerland_HG, and Natufian), modeling Steppe_EMBA and Baikal_LNBA.

Population	SampleID	ObservedSN	Representat	Lineage
Botai_CA	BOT15	N-M231	N-M231	N
Botai_CA	BOT14	R-M478	R-M478	R1b1a1
Kurma_EBA	DA354	Q-M1083	Q-L472	Q1a
Lokomotiv_EN	DA357	C-F4015	C-F4015	C2b1a1
Lokomotiv_EN	DA359	N-M2087.1	N-M2087.1	N1c
Okunevo_EMBA	RISE673	Q-M1100	Q-L472	Q1a
Okunevo_EMBA	RISE683	Q-L712	Q-L712	Q1a1b1
Okunevo_EMBA	RISE672	Q-M346	Q-M346	Q1a2
Okunevo_EMBA	RISE674	Q-M346	Q-M346	Q1a2
Okunevo_EMBA	RISE662	Q-L54	Q-L54	Q1a2a1
Okunevo_EMBA	RISE664	Q-L330	Q-L330	Q1a2a1c
Okunevo_EMBA	RISE718	Q-L330	Q-L330	Q1a2a1c
Okunevo_EMBA	RISE719	Q-L334	Q-L330	Q1a2a1c
Okunevo_EMBA	RISE670	Q-L940	Q-L940	Q1a2b
Okunevo_EMBA	RISE675	R-Z2105	R-Z2015	R1b1a2a2
Shamanka_EBA	DA334	Q-L55	Q-L53	Q1a2a
Shamanka_EBA	DA335	Q-L53	Q-L53	Q1a2a
Shamanka_EBA	DA336	Q-L53	Q-L53	Q1a2a
Shamanka_EBA	DA337	Q-L475	Q-L53	Q1a2a
Shamanka_EBA	DA338	Q-L53	Q-L53	Q1a2a
Shamanka_EBA	DA339	Q-L334	Q-L330	Q1a2a1c
Shamanka_EN	DA247	N-M231	N-M231	N
Shamanka_EN	DA251	N-M2291	N-M2291	N1
Shamanka_EN	DA245	N-L666	N-L666	N1c2
Shamanka_EN	DA248	N-L666	N-L666	N1c2
Shamanka_EN	DA362	N-L666	N-L666	N1c2
Shamanka_EN	DA250	NO-M214	NO-M214	NO1
UstIda_EBA	DA361	Q-M346	Q-M346	Q1a2
UstIda_EBA	DA353	Q-L476	Q-L53	Q1a2a
UstIda_EBA	DA356	Q-L213	Q-L53	Q1a2a
UstIda_EBA	DA343	Q-L54	Q-L54	Q1a2a1
UstIda_LN	DA345	N-M2080	N-M46	N1c1
UstIda_LN	DA355	Q-L892	Q-M346	Q1a2
Yamnaya	YamnayaKaragash_EMBA	R-CTS1843	R-CTS1843	R1b1a2a2c1
Turkmenistan_IA	DA382	R-F992	R-F992	R1a1a1b2
Namazga_CA	DA379	J-L134	J-M304	J
Namazga_CA	DA381	J-L26	J-L26	J2a1
Anatolia_EBA	MA2212	J-L559	J-M410	J2a
Anatolia_MLBA	MA2200	J-L26	J-L26	J2a1
Anatolia_MLBA	MA2205	J-L27	J-L26	J2a1
Anatolia_MLBA	MA2208	G-M3317	G-M406	G2a2b1

Table S13.

Y-chromosome lineages identified in 41 ancient males from the present study. Observed SNP is the marker for which at least 1 derived allele was identified in the data. Representative SNP is the marker that is deemed representative of the Observed SNP and may not have been directly genotyped.

Group	SampleID	Haplogroup	Quality	Group	SampleID	Haplogroup	Quality
Anatolia_EBA	MA2210	H	0.6623	Okunevo_EMBA	RISE675	D4+195	0.8728
Anatolia_EBA	MA2212	W5	0.7906	Okunevo_EMBA	RISE677	A8a1	0.808
Anatolia_EBA	MA2213	J1c10a	0.9048	Okunevo_EMBA	RISE680	A+152+16362	0.8055
Anatolia_IA	MA2197	U8b1b2	0.57	Okunevo_EMBA	RISE681	A8a1	0.8748
Anatolia_MLBA	MA2200	K1a+150	0.834	Okunevo_EMBA	RISE683	H15b1	0.7518
Anatolia_MLBA	MA2203	J1c	0.8591	Okunevo_EMBA	RISE684	C5c	0.8222
Anatolia_MLBA	MA2205	J2b1	0.7712	Okunevo_EMBA	RISE685	C5c	0.9046
Anatolia_MLBA	MA2206	U1a	0.8627	Okunevo_EMBA	RISE718	C5c	0.8503
Anatolia_MLBA	MA2208	H6a1b2e	0.5296	Okunevo_EMBA	RISE719	C5c	0.8824
Anatolian_Ottoman	MA2195	D4j	0.8464	Shamanka_EBA	DA334	C4a2a1	0.8686
Anatolian_Ottoman	MA2196	K	0.7099	Shamanka_EBA	DA335	F1b1b	0.877
Botai_CA	BOT14	K1b2	0.9639	Shamanka_EBA	DA336	C4a2a1	0.8877
Botai_CA	BOT15	R1b1	0.9265	Shamanka_EBA	DA337	C4a1a3	0.9003
Botai_CA	BOT2016	Z1a	0.9412	Shamanka_EBA	DA338	C4a2a1	0.9011
CentralSteppe_EMBA	EBA1	C4+152	0.9078	Shamanka_EBA	DA339	G2a1	0.8659
CentralSteppe_EMBA	EBA2	C4a1a4a	0.9483	Shamanka_EN	DA245	G2a1	0.8493
Kurma_EBA	DA354	D4	0.6069	Shamanka_EN	DA246	D4e1	0.9434
Kurma_EBA	DA358	F1b	0.8391	Shamanka_EN	DA247	C4	0.884
Kurma_EBA	DA360	F1b	0.791	Shamanka_EN	DA248	C4	0.9122
Lokomotiv_EN	DA340	D4	0.8631	Shamanka_EN	DA249	C4	0.8622
Lokomotiv_EN	DA341	D4j	0.9006	Shamanka_EN	DA250	G2a1	0.8354
Lokomotiv_EN	DA357	A+152+16362	0.7649	Shamanka_EN	DA251	D4j	0.8919
Lokomotiv_EN	DA359	D4+195	0.8569	Shamanka_EN	DA252	G2a1	0.8709
Namazga_CA	DA380	U2b	0.6851	Shamanka_EN	DA253	F1b1+@152	0.9005
Namazga_CA	DA381	J1+16193	0.8103	Shamanka_EN	DA362	D4e1	0.921
Namazga_CA	DA383	W3a2	0.7657	SidelkinoEHG_ML	Sidelkino	U5a2	0.8538
Okunevo_EMBA	RISE515	A8a	0.7831	Turkmenistan_IA	DA382	T2c1a	0.7975
Okunevo_EMBA	RISE516	H6a1b	0.9117	UstIda_EBA	DA343	D4j4	0.9308
Okunevo_EMBA	RISE662	H6a	0.8019	UstIda_EBA	DA353	H2a2a	0.6306
Okunevo_EMBA	RISE664	A8a1	0.8254	UstIda_EBA	DA356	C4a1a3	0.892
Okunevo_EMBA	RISE667	A8a	0.7565	UstIda_EBA	DA361	C4a1a3	0.8999
Okunevo_EMBA	RISE670	A8a	0.7831	UstIda_LN	DA342	R1b1	0.787
Okunevo_EMBA	RISE671	H6a1b	0.8647	UstIda_LN	DA344	A+152+16362	0.7901
Okunevo_EMBA	RISE672	H6a1b	0.8647	UstIda_LN	DA345	D4j	0.8996
Okunevo_EMBA	RISE673	A8a	0.7316	UstIda_LN	DA355	A2	0.8086
Okunevo_EMBA	RISE674	A+152+16362	0.7954	YamnayaKaragash_EBA	Yamnaya	R1a1a	0.9641

Table S15. Mitochondrial DNA lineages identified in 74 ancient samples sequenced in the present study with Haplogrep.

Model	nSNPs	*p*-value Rank = 0 (1 stream)	*p*-value Rank = 1 (2 streams)	p-value Rank = 2 (3 streams)
Namazga+Onge	83533	4.89E-35	1	-
SteppeMLBA+Onge	87093	2.08E-169	1	-
Namazga+Xiongnu	108875	5.16E-58	1	-
Xiongnu+Onge	102064	8.46E-37	1	-
ZarafshanIA+Xiongnu	107624	1.24E-42	1	-
IranN+Xiongnu	111478	1.36E-51	1	-
IranN+SteppeEMBA	78724	2.51E-35	1	-
IranN+CHG	127395	0.00011	1	-
IranN+EHG	121110	2.93E-65	1	-
Namazga+Paniya	69112	2.96E-25	1	-
Namazga+Onge+SteppeMLBA	68094	1.97E-152	3.28E-08	1
Namazga+Onge+Xiongnu	76376	4.56E-76	2.56E-20	1
Namazga+Xiongnu +SteppeMLBA	74198	3.18E-169	1.16E-12	1
IranN+SteppeEMBA+Onge	70986	7.14E-168	4.86E-15	1
IranN+Xiongnu+SteppeMLBA	78041	1.10E-184	5.62E-20	1
ZarafshanIA+Xiongnu+SteppeMLBA	72211	4.82E-150	0.008	1
IranN+SteppeMLBA+Onge	73290	7.42E-175	3.25E-10	1
IranN+EHG+Onge	79549	2.74E-143	1.78E-20	1
IranN+CHG+Onge	82839	3.82E-49	0.00026	1
EHG+CHG+IranN	101164	3.70E-76	0.018	1
Namazga+Paniya+SteppeMLBA	63333	2.02E-146	1.95E-07	1

Table S16.

qpWave results for assessing outgroup informativeness in qpAdm models using all sites.
For each of the models that we tested in qpAdm, we used qpWave to assess whether the ancestries of the source populations could be modeled as independent streams of migration from a set of seven outgroups (Ust_Ishim, Anzick1, Kostenki14, Switzerland_HG, Natufian, Mal'ta). In the table, we show the number of SNPs used for each comparison (nSNPs) and qpWave *p*-values for the tests for 1, 2, and 3 streams of migration. For this test, we rejected the null hypotheses in each column (number of streams), when we observed $p<0.05$.

Model	nSNPs	*p*-value Rank=0 (1 stream)	*p*-value Rank=1 (2 streams)	*p*-value Rank=2 (3 streams)
Namazga+Onge	14855	8.74E-10	1	-
SteppeMLBA+Onge	15592	4.28E-49	1	-
Namazga+Xiongnu	19871	7.57E-13	1	-
Xiongnu+Onge	18475	9.80E-10	1	-
TurkmenistanIA+Xiongnu	19688	5.58E-13	1	-
IranN+Xiongnu	20084	1.02E-13	1	-
IranN+SteppeEMBA	14038	1.21E-08	1	-
IranN+CHG	22789	0.0083	1	-
IranN+EHG	21605	7.45E-17	1	-
Namazga+Paniya	12197	6.65E-07	1	-
Namazga+Onge+SteppeMLBA	12072	1.30E-42	1.76E-01	1
Namazga+Onge+Xiongnu	13681	1.25E-18	1.40E-07	1
Namazga+Xiongnu+SteppeMLBA	13425	2.30E-39	7.01E-02	1
IranN+SteppeEMBA+Onge	12510	1.51E-40	1.54E-05	1
IranN+Xiongnu+SteppeMLBA	13997	5.65E-46	9.40E-07	1
Turkmenistan_IA+Xiongnu+SteppeMLBA	13073	1.52E-32	0.542058833	1
IranN+SteppeMLBA+Onge	12919	1.45E-44	1.62E-04	1
IranN+EHG+Onge	14016	1.52E-36	1.57E-07	1
IranN+CHG+Onge	14631	1.17E-13	0.00680204	1
EHG+CHG+IranN	18044	7.10E-18	0.045000504	1
Namazga+Paniya+SteppeMLBA	11185	1.67E-41	1.32E-01	1

Table S17.

qpWave results for assessing outgroup informativeness in qpAdm models using transversion polymorphisms only. This table is similar to Table S5, but only transversion polymorphisms were used in each test. While this table recapitulates the general trends in Table S5, we observed some inconsistencies in the *p*-values for some tests. We interpret these as reduced statistical power in the dataset where transition polymorphisms were excluded.

Table captions for separate tables

Table S3.

Information for the samples and archaeological sites analysed in the present-study.
Detailed information of radiocarbon dating, archaeological context, isotopes, and geographical location associated to the sites and samples here analyzed.

Table S14.

Ancestral and derived SNP count supporting Y-chromosome lineage determination.
We present the number of markers which informed Y-chromosome haplogroup determination in our male samples.

References

1. 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. Della 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. Mkrtchyan, 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. Szeverényi, G. Tóth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, 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). doi:10.1038/nature14507

2. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo 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). doi:10.1038/nature14317

3. P. B. Damgaard, N. Marchi, S. Rasmussen, M. Peyrot, G. Renaud, Th. Korneliussen, 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. Khashkhanov, 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. Moiyesev, G. Mukhtarova, B. Nurmukhanbetov, Zh. Orozbekova, I. Panyushkina, K. Pieta, V. Smrčka, I. Shevnina, A. Logvin, K.-G. Sjögren, T. Štolcová, K. Tashbaeva, A. Tkachev, T. Tulegenov, D. Voyakin, L. Yepiskoposyan, S. Undrakhbold, V. Varfolomeev, A. Weber, 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 steppe. *Nature* 10.1038/s41586-018-0094-2 (2018).

4. D. W. Anthony, *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2010).

5. A. K. Outram, in *The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers*, V. Cummings, P. Jordan, M. Zvelebil, Eds. (Oxford Univ. Press, Oxford, 2014), pp. 749–766.

6. J. P. Mallory, *In Search of the Indo-Europeans: Language, Archaeology and Myth* (Thames & Hudson, London, 1989).

7. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-

Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015). doi:10.1038/ncomms9912

8. J. P. Mallory, D. Q. Adams, *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world* (Oxford Univ. Press, Oxford, 2006).

9. K. Kristiansen, M. E. Allentoft, K. M. Frei, R. Iversen, N. N. Johannsen, G. Kroonen, Ł. Pospieszny, T. D. Price, S. Rasmussen, K.-G. Sjögren, M. Sikora, E. Willerslev, Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity* **91**, 334–347 (2017). doi:10.15184/aqy.2017.17

10. A. K. Outram, N. A. Stear, R. Bendrey, S. Olsen, A. Kasparov, V. Zaibert, N. Thorpe, R. P. Evershed, The earliest horse harnessing and milking. *Science* **323**, 1332–1335 (2009). doi:10.1126/science.1168594

11. E. Kaiser, Der Übergang zur Rinderzucht im nördlichen Schwarzmeerraum. *Godišnjak Centar za Balkanološka Ispitivanja* **39**, 23–34 (2010).

12. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* **461**, 489–494 (2009). doi:10.1038/nature08365

13. K. Kristiansen, *Europe Before History* (Cambridge Univ. Press, Cambridge, 1998).

14. See the supplementary materials.

15. A. K. Outram, A. Polyakov, A. Gromov, V. Moiseyev, A. W. Weber, V. I. Bazaliiskii, O. I. Goriunova, Supplementary discussion of the archaeology of Central Asian and East Asian Neolithic to Bronze Age hunter-gatherers and early pastoralists, including consideration of horse domestication. 10.5281/zenodo.1240521 (9 May 2018).

16. V. V. Evdokimov, V. G. Loman, Raskopki yamnogo kurgana v Karagandinskoj oblasti, in *Voprosy Arheologii Central'nogo i Severnogo Kazahstana* (Karaganda, 1989), pp. 34–46.

17. D. W. Anthony, D. R. Brown, The secondary products revolution, horse-riding, and mounted warfare. *J. World Prehist.* **24**, 131–160 (2011). doi:10.1007/s10963-011-9051-9

18. V. I. Bazaliiskii, in *Prehistoric Foragers of the Cis-Baikal, Siberia*, A. Weber, H. McKenzie, Eds. (Northern Hunter-Gatherers Research Series, Canadian Circumpolar Institute Press, Edmonton, 2003), vol. 1.

19. V. I. Bazaliiskii, in *Prehistoric Hunter-Gatherers of the Baikal Region, Siberia: Bioarchaeological Studies of Past Life Ways*, A. Weber, M. A. Katzenberg, T. G. Schurr, Eds. (University of Pennsylvania Press, Philadelphia, 2010).

20. A. Weber, Social evolution among neolithic and early bronze age foragers in the Lake Baikal region: New light on old models. *Arctic Anthropol.* **31**, 1–15 (1994).

21. A. Weber, The Neolithic and Early Bronze Age of the Lake Baikal Region: A review of recent research. *J. World Prehist.* **9**, 99–165 (1995). doi:10.1007/BF02221004

22. A. Weber, V. I. Baziliiskii, Mortuary practices and social relations among the Neolithic foragers of the Angara and Lake Baikal region: retrospection and prospection, in *Debating complexity: proceedings of the 26th annual conference of the Archaeological*

Association of the University of Calgary, D. A. Meyer, P. C. Dawson, D. T. Hanna, Eds. (University of Calgary, Calgary 1003), pp. 97– 103.

23. V. F. Zaibert, *Botaiskaya Kultura* (KazAkparat, Almaty, 2009).

24. S. V. Svyatko, J. P. Mallory, E. M. Murphy, A. V. Polyakov, P. J. Reimer, R. J. Schulting, New radiocarbon dates and a review of the chronology of prehistoric populations from the Minusinsk basin, southern Siberia, Russia. *Radiocarbon* **51**, 243–273 (2009). doi:10.1017/S0033822200033798

25. 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). doi:10.1126/sciadv.1601877

26. 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). doi:10.1038/nature12736

27. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016). doi:10.1038/nature17993

28. J. A. Kamm, J. Terhorst, Y. S. Song, Efficient computation of the joint sample frequency spectra for multiple populations. *J. Comput. Graph. Stat.* **26**, 182–194 (2017). doi:10.1080/10618600.2016.1159212

29. A.-M. Ilumäe, M. Reidla, M. Chukhryaeva, M. Järve, H. Post, M. Karmin, L. Saag, A. Agdzhoyan, A. Kushniarevich, S. Litvinov, N. Ekomasova, K. Tambets, E. Metspalu, R. Khusainova, B. Yunusbayev, E. K. Khusnutdinova, L. P. Osipova, S. Fedorova, O. Utevska, S. Koshel, E. Balanovska, D. M. Behar, O. Balanovsky, T. Kivisild, P. A. Underhill, R. Villems, S. Rootsi, Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *Am. J. Hum. Genet.* **99**, 163–173 (2016). doi:10.1016/j.ajhg.2016.05.025

30. V. T. Kovaleva, N. M. Chairkina, "Etnokul'turnye i etnogeneticheskie protsessy v severnem zaurale v kontse kamennogo–nachale bronzovogo veka: Itogi i problemy issledovaniya"

in *Voprosy arkheologii Urala* (Uralskii gosudarstvenyi universitet, Ekaterinburg, 1991), pp. 45–70.

31. D. W. Anthony, D. R. Brown, The secondary products revolution, horse-riding, and mounted warfare. *J. World Prehist.* **24**, 131–160 (2011). doi:10.1007/s10963-011-9051-9

32. C. Gaunitz, A. Fages, K. Hanghøj, A. Albrechtsen, N. Khan, M. Schubert, A. Seguin-Orlando, I. J. Owens, S. Felkel, O. Bignon-Lau, P. de Barros Damgaard, A. Mittnik, A. F. Mohaseb, H. Davoudi, S. Alquraishi, A. H. Alfarhan, K. A. S. Al-Rasheid, E. Crubézy, N. Benecke, S. Olsen, D. Brown, D. Anthony, K. Massy, V. Pitulko, A. Kasparov, G. Brem, M. Hofreiter, G. Mukhtarova, N. Baimukhanov, L. Lõugas, V. Onar, P. W. Stockhammer, J. Krause, B. Boldgiv, S. Undrakhbold, D. Erdenebaatar, S. Lepetz, M. Mashkour, A. Ludwig, B. Wallner, V. Merz, I. Merz, V. Zaibert, E. Willerslev, P. Librado, A. K. Outram, L. Orlando, Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–114 (2018). doi:10.1126/science.aao3297

33. S. L. Olsen, S. Grant, A. M. Choyke, L. Bartosiewicz, *Horses and Humans: the Evolution of Human-Equine Relationships* (British Archeological Reports, Oxford, 2006).

34. N. M. Myres, S. Rootsi, A. A. Lin, M. Järve, R. J. King, I. Kutuev, V. M. Cabrera, E. K. Khusnutdinova, A. Pshenichnov, B. Yunusbayev, O. Balanovsky, E. Balanovska, P. Rudan, M. Baldovic, R. J. Herrera, J. Chiaroni, J. Di Cristofaro, R. Villems, T. Kivisild, P. A. Underhill, A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* **19**, 95–101 (2011). doi:10.1038/ejhg.2010.146

35. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012). doi:10.1371/journal.pgen.1002453

36. S. Schiffels, W. Haak, P. Paajanen, B. Llamas, E. Popescu, L. Loe, R. Clarke, A. Lyons, R. Mortimer, D. Sayer, C. Tyler-Smith, A. Cooper, R. Durbin, Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016). doi:10.1038/ncomms10408

37. A. A. Kovalev, D. Erdenebaatar, "Discovery of new cultures of the Bronze Age in Mongolia according to the data obtained by the International Central Asian Archaeological Expedition" in *Current Archaeoligcal Research in Mongolia*, J. Bemmann, H. Parzinger, E. Pohl, D. Tseveendorzh, Eds. (Vor- und Frühgeschichtliche Archäologie Uni-Bonn, Bonn, 2009), pp. 149–170.

38. G. Larson, E. K. Karlsson, A. Perri, M. T. Webster, S. Y. W. Ho, J. Peters, P. W. Stahl, P. J. Piper, F. Lingaas, M. Fredholm, K. E. Comstock, J. F. Modiano, C. Schelling, A. I. Agoulnik, P. A. Leegwater, K. Dobney, J.-D. Vigne, C. Vilà, L. Andersson, K. Lindblad-Toh, Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8878–8883 (2012). doi:10.1073/pnas.1203005109

39. K. H. Røed, O. Flagstad, M. Nieminen, O. Holand, M. J. Dwyer, N. Røv, C. Vilà, Genetic analyses reveal independent domestication origins of Eurasian reindeer. *Proc. Biol. Sci.* **275**, 1849–1855 (2008). doi:10.1098/rspb.2008.0332

40. M. A. Zeder, Pathways to animal domestication. *Biodivers. Agric. Domest. Evol Sustain.* **2012**, 227–259 (2012).

41. R. Willerslev, P. Vitebsky, A. Alekseyev, Sacrifice as the ideal hunt: A cosmological explanation for the origin of reindeer domestication. *J. R. Anthropol. Inst.* **21**, 1–23 (2015). doi:10.1111/1467-9655.12142

42. 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-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, 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. Shamoon-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyan, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016). doi:10.1038/nature19310

43. A. Basu, N. Sarkar-Roy, P. P. Majumder, Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1594–1599 (2016). doi:10.1073/pnas.1513197113

44. H.-P. Francfort, L'âge du bronze en Asie centrale: La civilisation de l'Oxus. *Anthropol. Middle East.* **4**, 91–111 (2009).

45. E. E. Kuz′mina, *The Origin of the Indo-Iranians* (Brill, Leiden, 2007).

46. A. Parpola, *The Roots of Hinduism: The Early Aryans and the Indus Civilization* (Oxford Univ. Press, USA, 2015).

47. 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). doi:10.1038/nature16152

48. K. Kristiansen, B. Hemphill, G. Barjamovic, S. Omura, S. Y. Senyurt, V. Moiseyev, A. Gromov, F. E. Yediay, H. Ahmad, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, P. B. Damgaard, Archaeological supplement A to Damgaard et al. 2018: Archaeology of the Caucasus, Anatolia, Central and South Asia 4000-1500 BCE. 10.5281/zenodo.1240516 (9 May 2018).

49. G. Kroonen, G. Barjamovic, M. Peyrot, Linguistic supplement to Damgaard et al. 2018: Early Indo-European Languages, Anatolian, Tocharian and Indo-Iranian. 10.5281/zenodo.1240524 (9 May 2018).

50. K. Kristiansen, T. B. Larsson, *The Rise of Bronze Age Society: Travels, Transmissions and Transformations* (Cambridge Univ. Press, Cambridge, 2005).

51. M. Bonechi, Aleppo in età arcaica. A proposito di un'opera recente. *SEL* **7**, 15–37 (1990).

52. M. Mayrhofer, *Die Indo-Arier im Alten Vorderasien: Mit Einer Analytischen Bibliographie* (Harrassowitz, 1966).

53. H. C. Melchert, "The dialectal position of Anatolian within Indo-European" in *Annual Meeting of the Berkeley Linguistics Society* (1998), pp. 24–31.

54. J. Puhvel, "Whence the Hittite, whither the Jonesian vision?" in *Sprung from a common source*, S. M. Lamb, E. D. Mitchell, Eds. (Stanford Univ. Press, Stanford, Calif., 1991), pp. 51–66.

55. F. Kortlandt, C. C. Uhlenbeck on Indo-European, Uralic and Caucasian. *Hist. Sprachforsch* **122**, 39–47 (2009).

56. S. M. Winn, Burial evidence and the kurgan culture in Eastern Anatolia c. 3000 BC: An interpretation. *J. Indo-Eur. Stud.* **9**, 113–118 (1981).

57. I. Lazaridis, A. Mittnik, N. Patterson, S. Mallick, N. Rohland, S. Pfrengle, A. Furtwängler, A. Peltzer, C. Posth, A. Vasilakis, P. J. P. McGeorge, E. Konsolaki-Yannopoulou, G. Korres, H. Martlew, M. Michalodimitrakis, M. Özsait, N. Özsait, A. Papathanasiou, M. Richards, S. A. Roodenberg, Y. Tzedakis, R. Arnott, D. M. Fernandes, J. R. Hughey, D. M. Lotakis, P. A. Navas, Y. Maniatis, J. A. Stamatoyannopoulos, K. Stewardson, P. Stockhammer, R. Pinhasi, D. Reich, J. Krause, G. Stamatoyannopoulos, Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).

58. B. Darden, "On the question of the Anatolian origin of Indo-Hittite" in *Greater Anatolia and the Indo-Hittite Language Family*, R. Drews, Ed. (Journal of Indo-European Studies, Washington, 2001), pp. 184–228.

59. A. T. Smith, *The Political Machine: Assembling Sovereignty in the Bronze Age Caucasus* (Princeton Univ. Press, 2015).

60. A. Sagona, *The Archaeology of the Caucasus: From Earliest Settlements to the Iron Age* (Cambridge Univ. Press, 2017).

61. C. Burney, D. M. Lang, *The Peoples of the Hills: Ancient Ararat and Caucasus* (Weidenfeld and Nicolson, London, 1971).

62. M. T. Larsen, *Ancient Kanesh: A Merchant Colony in Bronze Age Anatolia* (Cambridge Univ. Press, 2015).

63. M. Forlanini, "An attempt at reconstructing the branches of the Hittite royal family of the Early Kingdom period" in *Pax Hethitica: Studies on the Hittites and their neighbours in honour of Itamar Singer*, Y. Cohen, A. Gilan, J. L. Miller, Eds. (Harrasowitz, Wiesbaden, 2010), p. 115.

64. P. Skoglund, H. Malmström, A. Omrak, M. Raghavan, C. Valdiosera, T. Günther, P. Hall, K. Tambets, J. Parik, K.-G. Sjögren, J. Apel, E. Willerslev, J. Storå, A. Götherström, M. Jakobsson, Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014). doi:10.1126/science.1253448

65. N. Lazaridis, A. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G.

Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi:10.1038/nature13673

66. L. Rahmstorf, "Indications of Aegean-Caucasian relations during the third millennium BC", in *Von Maikop bis Trialeti. Gewinnung und Verbreitung von Metallen und Obsidian in Kaukasien im 4.-2. Jt. v. Chr.*, S. Hansen, A. Hauptmann, I. Motzenbäcker, E. Pernicka, Eds. (Verlag Rudolf Habelt, Bonn, 2010), pp. 263–295.

67. I. Mathieson, S. Alpaslan-Roodenberg, C. Posth, A. Szécsényi-Nagy, N. Rohland, S. Mallick, I. Olalde, N. Broomandkhoshbacht, F. Candilio, O. Cheronet, D. Fernandes, M. Ferry, B. Gamarra, G. G. Fortes, W. Haak, E. Harney, E. Jones, D. Keating, B. Krause-Kyora, I. Kucukkalipci, M. Michel, A. Mittnik, K. Nägele, M. Novak, J. Oppenheimer, N. Patterson, S. Pfrengle, K. Sirak, K. Stewardson, S. Vai, S. Alexandrov, K. W. Alt, R. Andreescu, D. Antonović, A. Ash, N. Atanassova, K. Bacvarov, M. B. Gusztáv, H. Bocherens, M. Bolus, A. Boroneanț, Y. Boyadzhiev, A. Budnik, J. Burmaz, S. Chohadzhiev, N. J. Conard, R. Cottiaux, M. Čuka, C. Cupillard, D. G. Drucker, N. Elenski, M. Francken, B. Galabova, G. Ganetsovski, B. Gély, T. Hajdu, V. Handzhyiska, K. Harvati, T. Higham, S. Iliev, I. Janković, I. Karavanić, D. J. Kennett, D. Komšo, A. Kozak, D. Labuda, M. Lari, C. Lazar, M. Leppek, K. Leshtakov, D. L. Vetro, D. Los, I. Lozanov, M. Malina, F. Martini, K. McSweeney, H. Meller, M. Menđušić, P. Mirea, V. Moiseyev, V. Petrova, T. D. Price, A. Simalcsik, L. Sineo, M. Šlaus, V. Slavchev, P. Stanev, A. Starović, T. Szeniczey, S. Talamo, M. Teschler-Nicola, C. Thevenet, I. Valchev, F. Valentin, S. Vasilyev, F. Veljanovska, S. Venelinova, E. Veselovskaya, B. Viola, C. Virag, J. Zaninović, S. Zäuner, P. W. Stockhammer, G. Catalano, R. Krauß, D. Caramelli, G. Zariņa, B. Gaydarska, M. Lillie, A. G. Nikitin, I. Potekhina, A. Papathanasiou, D. Borić, C. Bonsall, J. Krause, R. Pinhasi, D. Reich, The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018). doi:10.1038/nature25778

68. S. Lumsden, "Material culture and the Middle Ground in the Old Assyrian Colony period" in *Old Assyrian studies in Memory of Paul Garelli* (Nederlands Instituut voor het Nabije Oosten, 2008), pp. 21–43.

69. M. T. Larsen, A. W. Lassen, "Cultural exchange at Kültepe" in *Extraction and Control: Studies In Honor of Matthew W Stolper* (Oriental Institute of the University of Chicago, 2014), pp. 171–188.

70. A. Lehrman, “Reconstructing Proto-Indo-Hittite” in *Greater Anatolia and the Indo-Hittite Language Family*, R. Drews, Ed. (Journal of Indo-European Studies monographs, 2001), vol. 38, pp. 106–130.

71. C. Watkins, “An Indo-European linguistic area and its characteristics: Anatolia” in *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, A.Y. Aikhenvald, R. M. W. Dixon, Eds. (Oxford Univ. Press, 2001), pp. 44–63.

72. N. Oettinger, “Indogermanische Sprachträger lebten schon im 3. Jahrtausend v. Chr. in Kleinasien” in *Die Hethiter und ihr Reich: Das Volk der 1000 Götter*, T. Özgüç, Ed. (Stuttgart, Theiss, 2002), pp. 50–55.

73. H. C. Melchert, Ed., *The Luwians* (Brill, 2003).

74. J. McKinley, “Compiling a skeletal inventory: disarticulated and co-mingled remains” in *Guidelines to the Standards for Recording Human Remains*, M. Brickley, J. McKinley, Eds. (BABAO and IFA, Reading, 2004) pp. 14–17.

75. J. H. Schwartz, *Skeleton Keys: An Introduction to Human Skeletal Morphology, Development and Analysis* (Oxford Univ. Press, Oxford, 2007).

76. T. W. Phenice, A newly developed visual method of sexing the os pubis. *Am. J. Phys. Anthropol.* **30**, 297–301 (1969). doi:10.1002/ajpa.1330300214

77. D. Ferembach, I. Schwindezky, M. Stoukal, Recommendations for age and sex diagnoses of skeletons. *J. Hum. Evol.* **9**, 517–549 (1980). doi:10.1016/0047-2484(80)90061-5

78. S. R. Loth, M. Henneberg, Mandibular ramus flexure: A morphologic indicator of sexual dimorphism in the human skeleton. *Am. J. Phys. Anthropol.* **99**, 473–485 (1996). doi:10.1002/(SICI)1096-8644(199603)99:3<473:AID-AJPA8>3.0.CO;2-X

79. S. T. Brooks, J. M. Suchey, Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol.* **5**, 227–238 (1990). doi:10.1007/BF02437238

80. C. O. Lovejoy, R. S. Meindl, T. R. Pryzbeck, R. P. Mensforth, Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *Am. J. Phys. Anthropol.* **68**, 15–28 (1985). doi:10.1002/ajpa.1330680103

81. D. R. Brothwell, *Digging up Bones* (Oxford Univ. Press, Oxford, 1981).

82. M. Trotter, “Estimation of stature from intact long bones” in *Personal Identification in Mass Disasters*, T. D. Stewart, Ed. (Smithsonian Institution, Washington, 1970), pp. 71–83.

83. I. V. Mertz, V. K. Mertz, “Novye materialy rannego bronzovogo veka iz Zapadnoj chasti Kulundinskoj ravniny” in *Sohranenie i izuchenie kul'turnogo nasledija Altajskogo kraja: Materialy XVIII XIX regional'nyh nauchno-prakticheskih konferencij*, G. A. Kubareva, V. P. Semibratov, Eds. (Azbuka, Barnaul, 2013), pp. 207–215.

84. V. K. Merz, I. V. Merz Pogrebeniya, “‘yamnogo’ tipa Vostochnogo i Severo-Vostochnogo Kazakhstana (k postanovke problemy)” in *Afanasyevskiy Sbornik*, N. F. Stepanova, A. V. Polyakov, S. S. Tur, P. I. Shulga, Eds. (Azbuka, Barnaul, 2010), pp. 134–144.

85. G. A. Maksimenkov, “Okunevskaya Kultura,” thesis abstract, Institute of History, Philology and Philosophy, Novosibirsk (1975).

86. I. P. Lazaretov, “K voprosu o yamno-katakombnyh svyazyah okunevskoy kul’tury” in *Problemy izucheniya okunevskoy kul’tury*, D. G. Savinov, Ed. (St. Petersburg State University, St. Petersburg, 1995) pp. 14–16.

87. M. N. Komarova, Burials of the Okunev Ulus. *Sovetskaya Archaeologia* **9**, 47–60 (1947).

88. A. W. Weber, R. J. Schulting, C. B. Ramsey, V. I. Bazaliiskii, O. I. Goriunova, E. B. Natal’ia, Chronology of middle Holocene hunter–gatherers in the Cis-Baikal region of Siberia: Corrections based on examination of the freshwater reservoir effect. *Quat. Int.* **419**, 74–98 (2016). doi:10.1016/j.quaint.2015.12.003

89. N. M. Moussa, V. I. Bazaliiskii, O. I. Goriunova, F. Bamforth, A. W. Weber, Y-chromosomal DNA analyzed for four prehistoric cemeteries from Cis-Baikal, Siberia. *J. Arch. Sci. Rep.* **17**, 932–942 (2018).

90. M. P. Ovchinnikov, *Materialy dlia izucheniia drevnostei v okrestnostiakh Irkutska. Izvestia VSORGO Vol XXXV* (Irkutskaia tipolitografiia, Irkutsk, 1904).

91. V. I. Bazaliiskii, N. A. Savelyev, The wolf of Baikal: The “Lokomotiv” early Neolithic cemetery in Siberia (Russia). *Antiquity* **77**, 20–30 (2003). doi:10.1017/S0003598X00061317

92. A. P. Okladnikov, A. K. Konopatskiy, Hunters for seal on the Baikal Lake in the Stone and Bronze Ages. *Folk Dan. Ethnogr. Tidsskr.* **16**, 299–308 (1974).

93. A. A. Tiutrin, V. I. Bazaliiskii, Mogil’nik v ust’e reki Idy v doline Angary. *Arkheologiia Paleoekologiia Ehnologiia Sib Dalnego Vostoka* **1**, 85–90 (1996) [A cemetery at the mouth of the Ida River in the Angara River Valley].

94. A. W. Weber, H. G. McKenzie, A. R. Lieverse, O. I. Goriunova, Eds., *Kurma XI, a Middle Holocene hunter-gatherer cemetery on Lake Baikal, Siberia: Archaeological and Osteological Materials* (German Archaeological Institute, Berlin, 2012).

95. C. B. Ramsey, R. Schulting, O. I. Goriunova, V. I. Bazaliiskii, A. W. Weber, Analyzing radiocarbon reservoir offsets through stable nitrogen isotopes and Bayesian modeling: A case study using paired human and faunal remains from the Cis-Baikal region, Siberia. *Radiocarbon* **56**, 789–799 (2014). doi:10.2458/56.17160

96. T. Nomokonova, R. J. Losey, O. I. Goriunova, A. W. Weber, A freshwater old carbon offset in Lake Baikal, Siberia and problems with the radiocarbon dating of archaeological sediments: Evidence from the Sagan-Zaba II site. *Quat. Int.* **290**, 110–125 (2013). doi:10.1016/j.quaint.2012.06.007

97. R. J. Schulting, C. B. Ramsey, V. I. Bazaliiskii, O. I. Goriunova, A. W. Weber, Freshwater reservoir offsets investigated through paired human-faunal 14 C dating and stable carbon and nitrogen isotope analysis at Lake Baikal, Siberia. *Radiocarbon* **56**, 991–1008 (2014). doi:10.2458/56.17963

98. S. Omura, “Preliminary report on the ninth excavation at Kaman-Kalehöyük (1994)” in *Essays on Ancient Anatolia and Syria in the second and third millennium B.C.*, T. Mikasa, Ed. (Harrassowitz, Wiesbaden, 1996), pp. 87–163.

99. H. Hongo, “Continuity or changes: faunal remains from stratum IId at Kaman-Kalehöyük” in *Identifying Changes: The Transition from Bronze to Iron Ages in Anatolia and its*

Neighbouring Regions, B. Fischer, H. Genz, É. Jean, K. Köroğlu, Eds. (Türk Eskiçağ Bilimleri Enstitüsü, Istanbul, 2003), pp. 257–269.

100. S. Omura, "Kaman-Kalehöyük excavations in central Anatolia" in *The Oxford Handbook of Ancient Anatolia*, G. McMahon, S. Steadman, Eds. (Oxford Univ. Press, Oxford, 2011), pp. 1095–1111.

101. S. Y. Şenyurt, A. Akcay, Y. Kamış, "Ovaören 2013 Yılı Kazıları" in *36. Kazı Sonuçları Toplantısı,* H. Dönmez, Ed. (T. C. Kültür ve turizm bakanlığı, Ankara, 2014), vol. II, pp. 101–120.

102. S. Y. Şenyurt, A. Akçay, Y. Kamış, Nevşehir-Ovaören Kazısı 2012. *TEBE-Haberler* **36**, 24–26 (2013).

103. T. Şener, "Ovaören/Topakhöyük ve Teras Yerleşimi Erken Tunç Çağı Mezarlarının Arkeolojik ve Antropolojik Açıdan Değerlendirilmesi" thesis, Gazi University (2014).

104. V. M. Masson, G. A. Pugachenkova, *Jeitun settlement (problems of the origin of the producing economy)* (Nauka, Moscow, 1971).

105. B. A. Kuftin, Studies of the Southern Turkmenistan Archeological Complex Expedition on studies of the Anau culture in 1952. *Izvestiya Akademii Nauk Turkmenskoy SSR. Seriya Oschestvennye Nauki* **1**, 22–29 (1954).

106. B. A. Kuftin, Report on excavations of the XIV excavation team of the Southern Turkmenistan Archeological Complex Expedition of agricultural settlements of the Eneolithic and Bronze Age in 1952. *Trudy Uzhno-Turkmenistanskoy arkheologicheskoy kompleksnoy expeditsii.* **7**, 260–290 (1956).

107. N. Y. Merpert, "The Eneolithic of the Southern part of the USSR and Eurasian steppes" in *Eneolithic of the USSR*, V. M. Mason, N. Y. Merpert, Eds. (Nauka, Moscow, 1982), pp. 322–331.

108. V. M. Masson, *Srednyaya Asiya i Drevniy Vostok* (Nauka, Moscow-Leningrad, 1964).

109. N. A. Avanesova, D. M. Jurakulova, "The ancient nomads of Zarafshan" in *Culture of nomadic peoples of Central Asia* (Proceedings of the International Conference, Samarkand, MICAI, 2008), pp. 13–33.

110. P. B. Damgaard, A. Margaryan, H. Schroeder, L. Orlando, E. Willerslev, M. E. Allentoft, Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015). doi:10.1038/srep11184

111. E. Willerslev, A. Cooper, Review Paper. Ancient DNA. *Proc. R. Soc. London Ser. B* **272**, 3–16 (2005). doi:10.1098/rspb.2004.2813

112. L. Orlando, A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, B. Petersen, I. Moltke, P. L. F. Johnson, M. Fumagalli, J. T. Vilstrup, M. Raghavan, T. Korneliussen, A.-S. Malaspinas, J. Vogt, D. Szklarczyk, C. D. Kelstrup, J. Vinther, A. Dolocan, J. Stenderup, A. M. V. Velazquez, J. Cahill, M. Rasmussen, X. Wang, J. Min, G. D. Zazula, A. Seguin-Orlando, C. Mortensen, K. Magnussen, J. F. Thompson, J. Weinstock, K. Gregersen, K. H. Røed, V. Eisenmann, C. J. Rubin, D. C. Miller, D. F. Antczak, M. F. Bertelsen, S. Brunak, K. A. S. Al-Rasheid, O. Ryder, L. Andersson, J. Mundy, A. Krogh, M. T. P. Gilbert, K. Kjær, T. Sicheritz-Ponten, L. J.

Jensen, J. V. Olsen, M. Hofreiter, R. Nielsen, B. Shapiro, J. Wang, E. Willerslev, Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013). doi:10.1038/nature12323

113. R. Sinha, G. Stanley, G. S. Gulati, C. Ezran, K. J. Travaglini, Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. bioRxiv, 125724 [Preprint] (9 April 2017). https://doi.org/10.1101/125724

114. J. Dabney, M. Meyer, Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012). doi:10.2144/000113809

115. E. H. M. Wong, A. Khrunin, L. Nichols, D. Pushkarev, D. Khokhrin, D. Verbenko, O. Evgrafov, J. Knowles, J. Novembre, S. Limborska, A. Valouev, Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res.* **27**, 1–14 (2017). doi:10.1101/gr.202945.115

116. 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). doi:10.1093/bioinformatics/btp352

117. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016). doi:10.1186/s13104-016-1900-2

118. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN] (16 March 2013).

119. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011). doi:10.1038/ng.806

120. F. García-Alcalde, K. Okonechnikov, J. Carbonell, L. M. Cruz, S. Götz, S. Tarazona, J. Dopazo, T. F. Meyer, A. Conesa, Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678–2679 (2012). doi:10.1093/bioinformatics/bts503

121. Q. Fu, A. Mittnik, P. L. F. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, R. Schmitz, J. Burger, A. M. Ronchitelli, F. Martini, R. G. Cremonesi, J. Svoboda, P. Bauer, D. Caramelli, S. Castellano, D. Reich, S. Pääbo, J. Krause, A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). doi:10.1016/j.cub.2013.02.044

122. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E.

E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010). doi:10.1126/science.1188021

123. M. Rasmussen, X. Guo, Y. Wang, K. E. Lohmueller, S. Rasmussen, A. Albrechtsen, L. Skotte, S. Lindgreen, M. Metspalu, T. Jombart, T. Kivisild, W. Zhai, A. Eriksson, A. Manica, L. Orlando, F. M. De La Vega, S. Tridico, E. Metspalu, K. Nielsen, M. C. Ávila-Arcos, J. V. Moreno-Mayar, C. Muller, J. Dortch, M. T. P. Gilbert, O. Lund, A. Wesolowska, M. Karmin, L. A. Weinert, B. Wang, J. Li, S. Tai, F. Xiao, T. Hanihara, G. van Driem, A. R. Jha, F.-X. Ricaut, P. de Knijff, A. B. Migliano, I. Gallego Romero, K. Kristiansen, D. M. Lambert, S. Brunak, P. Forster, B. Brinkmann, O. Nehlich, M. Bunce, M. Richards, R. Gupta, C. D. Bustamante, A. Krogh, R. A. Foley, M. M. Lahr, F. Balloux, T. Sicheritz-Pontén, R. Villems, R. Nielsen, J. Wang, E. Willerslev, An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011). doi:10.1126/science.1211177

124. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013). doi:10.1016/j.jas.2013.07.004

125. M. Lipatov, K. Sanjeev, R. Patro, K. Veeramah, Maximum likelihood estimation of biological relatedness from low coverage sequencing data. bioRxiv, 023374 [Preprint] (29 July 2015). https://doi.org/10.1101/023374

126. J. E. Wigginton, D. J. Cutler, G. R. Abecasis, A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **76**, 887–893 (2005). doi:10.1086/429864

127. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi:10.1086/519795

128. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. B. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014). doi:10.1038/nature13810

129. O. Delaneau, J. Marchini, J.-F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011). doi:10.1038/nmeth.1785

130. S. Leslie, B. Winney, G. Hellenthal, D. Davison, A. Boumertit, T. Day, K. Hutnik, E. C. Royrvik, B. Cunliffe, D. J. Lawson, D. Falush, C. Freeman, M. Pirinen, S. Myers, M. Robinson, P. Donnelly, W. Bodmer, Wellcome Trust Case Control Consortium 2, International Multiple Sclerosis Genetics Consortium, The fine-scale genetic structure of the British population. *Nature* **519**, 309–314 (2015). doi:10.1038/nature14230

131. P. F. Palamara, L. C. Francioli, P. R. Wilton, G. Genovese, A. Gusev, H. K. Finucane, S. Sankararaman, S. R. Sunyaev, P. I. W. de Bakker, J. Wakeley, I. Pe'er, A. L. Price, Genome of the Netherlands Consortium, Leveraging Distant Relatedness to Quantify

Human Mutation and Gene-Conversion Rates. *Am. J. Hum. Genet.* **97**, 775–789 (2015). doi:10.1016/j.ajhg.2015.10.006

132. M. Lipson, P.-R. Loh, S. Sankararaman, N. Patterson, B. Berger, D. Reich, Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. *PLOS Genet.* **11**, e1005550 (2015). doi:10.1371/journal.pgen.1005550

133. 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). doi:10.1534/genetics.112.145037

134. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011). doi:10.1093/bioinformatics/btr509

135. G. D. Poznik, B. M. Henn, M.-C. Yee, E. Sliwerska, G. M. Euskirchen, A. A. Lin, M. Snyder, L. Quintana-Murci, J. M. Kidd, P. A. Underhill, C. D. Bustamante, Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562–565 (2013). doi:10.1126/science.1237619

136. G. D. Poznik, Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. bioRxiv, 088716 [Preprint]. (19 November 2016). https://doi.org/10.1101/088716

137. M. Rasmussen, S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford Jr., S. Rasmussen, I. Moltke, A. Albrechtsen, S. M. Doyle, G. D. Poznik, V. Gudmundsdottir, R. Yadav, A.-S. Malaspinas, S. S. White 5th, M. E. Allentoft, O. E. Cornejo, K. Tambets, A. Eriksson, P. D. Heintzman, M. Karmin, T. S. Korneliussen, D. J. Meltzer, T. L. Pierre, J. Stenderup, L. Saag, V. M. Warmuth, M. C. Lopes, R. S. Malhi, S. Brunak, T. Sicheritz-Ponten, I. Barnes, M. Collins, L. Orlando, F. Balloux, A. Manica, R. Gupta, M. Metspalu, C. D. Bustamante, M. Jakobsson, R. Nielsen, E. Willerslev, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014). doi:10.1038/nature13025

138. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. P. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Grønnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. G. Higham, C. B. Ramsey, T. V. O. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Pontén, R. Villems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010). doi:10.1038/nature08835

139. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014). doi:10.1038/ncomms6257

140. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). doi:10.1093/nar/gkh340

141. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016). doi:10.1093/molbev/msw054

142. M. Rasmussen, M. Sikora, A. Albrechtsen, T. S. Korneliussen, J. V. Moreno-Mayar, G. D. Poznik, C. P. E. Zollikofer, M. P. de León, M. E. Allentoft, I. Moltke, H. Jónsson, C. Valdiosera, R. S. Malhi, L. Orlando, C. D. Bustamante, T. W. Stafford Jr., D. J. Meltzer, R. Nielsen, E. Willerslev, The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).

143. M. C. Dulik, S. I. Zhadanov, L. P. Osipova, A. Askapuli, L. Gau, O. Gokcumen, S. Rubinstein, T. G. Schurr, Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* **90**, 229–246 (2012). doi:10.1016/j.ajhg.2011.12.014

144. T. Kivisild, The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* **136**, 529–546 (2017). doi:10.1007/s00439-017-1773-z

145. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi:10.1038/nature14558

146. C. Keyser, C. Hollard, A. Gonzalez, J.-L. Fausser, E. Rivals, A. N. Alexeev, A. Riberon, E. Crubézy, B. Ludes, The ancient Yakuts: A population genetic enigma. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130385 (2015). doi:10.1098/rstb.2013.0385

147. M. Karmin, L. Saag, M. Vicente, M. A. Wilson Sayres, M. Järve, U. G. Talas, S. Rootsi, A.-M. Ilumäe, R. Mägi, M. Mitt, L. Pagani, T. Puurand, Z. Faltyskova, F. Clemente, A. Cardona, E. Metspalu, H. Sahakyan, B. Yunusbayev, G. Hudjashov, M. DeGiorgio, E.-L. Loogväli, C. Eichstaedt, M. Eelmets, G. Chaubey, K. Tambets, S. Litvinov, M. Mormina, Y. Xue, Q. Ayub, G. Zoraqi, T. S. Korneliussen, F. Akhatova, J. Lachance, S. Tishkoff, K. Momynaliev, F.-X. Ricaut, P. Kusuma, H. Razafindrazaka, D. Pierron, M. P. Cox, G. N. N. Sultana, R. Willerslev, C. Muller, M. Westaway, D. Lambert, V. Skaro, L. Kovačevic, S. Turdikulova, D. Dalimova, R. Khusainova, N. Trofimova, V. Akhmetova, I. Khidiyatova, D. V. Lichman, J. Isakova, E. Pocheshkhova, Z. Sabitov, N. A. Barashkov, P. Nymadawa, E. Mihailov, J. W. T. Seng, I. Evseeva, A. B. Migliano, S. Abdullah, G. Andriadze, D. Primorac, L. Atramentova, O. Utevska, L. Yepiskoposyan, D. Marjanovic, A. Kushniarevich, D. M. Behar, C. Gilissen, L. Vissers, J. A. Veltman, E. Balanovska, M. Derenko, B. Malyarchuk, A. Metspalu, S. Fedorova, A. Eriksson, A. Manica, F. L. Mendez, T. M. Karafet, K. R. Veeramah, N. Bradman, M. F. Hammer, L. P. Osipova, O. Balanovsky, E. K. Khusnutdinova, K. Johnsen, M. Remm, M. G. Thomas, C. Tyler-Smith, P. A. Underhill, E. Willerslev, R. Nielsen, M. Metspalu, R. Villems, T. Kivisild, A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **25**, 459–466 (2015). doi:10.1101/gr.186684.114

148. P. A. Underhill, G. D. Poznik, S. Rootsi, M. Järve, A. A. Lin, J. Wang, B. Passarelli, J. Kanbar, N. M. Myres, R. J. King, J. Di Cristofaro, H. Sahakyan, D. M. Behar, A.

Kushniarevich, J. Šarac, T. Šaric, P. Rudan, A. K. Pathak, G. Chaubey, V. Grugni, O. Semino, L. Yepiskoposyan, A. Bahmanimehr, S. Farjadian, O. Balanovsky, E. K. Khusnutdinova, R. J. Herrera, J. Chiaroni, C. D. Bustamante, S. R. Quake, T. Kivisild, R. Villems, The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* **23**, 1–8 (2014).

149. Z. Hofmanová, S. Kreutzer, G. Hellenthal, C. Sell, Y. Diekmann, D. Díez-Del-Molino, L. van Dorp, S. López, A. Kousathanas, V. Link, K. Kirsanow, L. M. Cassidy, R. Martiniano, M. Strobel, A. Scheu, K. Kotsakis, P. Halstead, S. Triantaphyllou, N. Kyparissi-Apostolika, D. Urem-Kotsou, C. Ziota, F. Adaktylou, S. Gopalan, D. M. Bobo, L. Winkelbach, J. Blöcher, M. Unterländer, C. Leuenberger, Ç. Çilingiroğlu, B. Horejs, F. Gerritsen, S. J. Shennan, D. G. Bradley, M. Currat, K. R. Veeramah, D. Wegmann, M. G. Thomas, C. Papageorgopoulou, J. Burger, Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6886–6891 (2016). doi:10.1073/pnas.1523951113

150. V. Battaglia, V. Grugni, U. A. Perego, N. Angerhofer, J. E. Gomez-Palmieri, S. R. Woodward, A. Achilli, N. Myres, A. Torroni, O. Semino, The first peopling of South America: New evidence from Y-chromosome haplogroup Q. *PLOS ONE* **8**, e71390 (2013). doi:10.1371/journal.pone.0071390

151. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44** (W1), W58–W63 (2016). doi:10.1093/nar/gkw233

152. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi:10.1093/bioinformatics/btu033

153. S. A. Fedorova, M. Reidla, E. Metspalu, M. Metspalu, S. Rootsi, K. Tambets, N. Trofimova, S. I. Zhadanov, B. Hooshiar Kashani, A. Olivieri, M. I. Voevoda, L. P. Osipova, F. A. Platonov, M. I. Tomsky, E. K. Khusnutdinova, A. Torroni, R. Villems, Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): Implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* **13**, 127 (2013). doi:10.1186/1471-2148-13-127

154. P. Flegontov, N. E. Altinisik, P. Changmai, N. Rohland, Paleo-Eskimo genetic legacy across North America. bioRxiv, 203018 [Preprint]. (13 October 2017). https://doi.org/10.1101/203018

155. A. T. Duggan, M. Whitten, V. Wiebe, M. Crawford, A. Butthof, V. Spitsyn, S. Makarov, I. Novgorodov, V. Osakovsky, B. Pakendorf, Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal. *PLOS ONE* **8**, e83570 (2013). doi:10.1371/journal.pone.0083570

156. C. Der Sarkissian, O. Balanovsky, G. Brandt, V. Khartanovich, A. Buzhilova, S. Koshel, V. Zaporozhchenko, D. Gronenborn, V. Moiseyev, E. Kolpakov, V. Shumkin, K. W. Alt, E. Balanovska, A. Cooper, W. Haak, Genographic Consortium, Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLOS Genet.* **9**, e1003296 (2013). doi:10.1371/journal.pgen.1003296

157. M. Ingman, U. Gyllensten, A recent genetic link between Sami and the Volga-Ural region of Russia. *Eur. J. Hum. Genet.* **15**, 115–120 (2007). doi:10.1038/sj.ejhg.5201712

158. M. Derenko, B. Malyarchuk, T. Grzybowski, G. Denisova, U. Rogalla, M. Perkova, I. Dambueva, I. Zakharov, Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLOS ONE* **5**, e15214 (2010). doi:10.1371/journal.pone.0015214

159. D. Comas, S. Plaza, R. S. Wells, N. Yuldaseva, O. Lao, F. Calafell, J. Bertranpetit, Admixture, migrations, and dispersals in Central Asia: Evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* **12**, 495–504 (2004). doi:10.1038/sj.ejhg.5201160

160. E. Crubézy, S. Amory, C. Keyser, C. Bouakaze, M. Bodner, M. Gibert, A. Röck, W. Parson, A. Alexeev, B. Ludes, Human evolution in Siberia: From frozen bodies to ancient DNA. *BMC Evol. Biol.* **10**, 25 (2010). doi:10.1186/1471-2148-10-25

161. M. Pala, A. Olivieri, A. Achilli, M. Accetturo, E. Metspalu, M. Reidla, E. Tamm, M. Karmin, T. Reisberg, B. Hooshiar Kashani, U. A. Perego, V. Carossa, F. Gandini, J. B. Pereira, P. Soares, N. Angerhofer, S. Rychkov, N. Al-Zahery, V. Carelli, M. H. Sanati, M. Houshmand, J. Hatina, V. Macaulay, L. Pereira, S. R. Woodward, W. Davies, C. Gamble, D. Baird, O. Semino, R. Villems, A. Torroni, M. B. Richards, Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am. J. Hum. Genet.* **90**, 915–924 (2012). doi:10.1016/j.ajhg.2012.04.003

162. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi:10.1038/nature18964

163. I. Mathieson, G. McVean, Demography and the age of rare variants. *PLOS Genet.* **10**, e1004528 (2014). doi:10.1371/journal.pgen.1004528

164. T. D. O'Connor, W. Fu, NHLBI GO Exome Sequencing Project, ESP Population Genetics and Statistical Analysis Working Group, Emily Turner, J. C. Mychaleckyj, B. Logsdon, P. Auer, C. S. Carlson, S. M. Leal, J. D. Smith, M. J. Rieder, M. J. Bamshad, D. A. Nickerson, J. M. Akey, Rare variation facilitates inferences of fine-scale population structure in humans. *Mol. Biol. Evol.* **32**, 653–660 (2015).

165. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* **7**, e34131 (2012). doi:10.1371/journal.pone.0034131

166. G. N. Lance, W. T. Williams, Computer programs for hierarchical polythetic classification ("Similarity analyses"). *Comput. J.* **9**, 60–64 (1966). doi:10.1093/comjnl/9.1.60

167. M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, Cluster: luster analysis basics and extensions, R Package (2012).

168. P. Reimer, S. Hoper, J. MacDonald, R. Reimer, S. Svyatko, M. Thompson, "The Queen's University, Belfast: Laboratory Protocols Used for AMS Radiocarbon Dating at the 14CHRONO Centre", (Rep. 5/2015, Historic England, 2015).