The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia

Authors: Peter de Barros Damgaard[1]†, Rui Martiniano[2,3]†, Jack Kamm[2]†, J. Víctor Moreno-Mayar[1]†, Guus Kroonen[4,5], Michaël Peyrot[5], Gojko Barjamovic[6], Simon Rasmussen[7], Claus Zacho[1], Nurbol Baimukhanov[8], Victor Zaibert[9], Victor Merz[10], Arjun Biddanda[11], Ilja Merz[10], Valeriy Loman[12], Valeriy Evdokimov[12], Emma Usmanova[12], Brian Hemphill[13], Andaine Seguin-Orlando[1], Fulya Eylem Yediay[14], Inam Ullah[1,15], Karl-Göran Sjögren[16], Katrine Højholt Iversen[7], Jeremy Choin[1], Constanza de la Fuente[1], Melissa Ilardo[1], Hannes Schroeder[1], Vyacheslav Moiseyev[17], Andrey Gromov[17], Andrei Polyakov[18], Sachihiro Omura[19], Süleyman Yücel Senyurt[20], Habib Ahmad[15,21], Catriona McKenzie[22], Ashot Margaryan[1], Abdul Hameed[23], Abdul Samad[24], Nazish Gul[15], Muhammad Hassan Khokhar[25], O. I. Goriunova[26,27], Vladimir I. Bazaliiskii[27], John Novembre[10,28], Andrzej W. Weber[29], Ludovic Orlando[1,30], Morten E. Allentoft[1], Rasmus Nielsen[31], Kristian Kristiansen[16], Martin Sikora[1], Alan K. Outram[22], Richard Durbin[2,3]*, Eske Willerslev[1,2,32]*.

Affiliations:

[1] Centre for GeoGenetics, Natural History Museum, University of Copenhagen.

[2] Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK.

[3] Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

[4] Department of Nordic Studies and Linguistics, University of Copenhagen, Denmark.

[5] Leiden University Centre for Linguistics, Leiden University, The Netherlands.

[6] Department of Near Eastern Languages and Civilizations, Harvard University, USA.

[7] Department of Bio and Health Informatics, Technical University of Denmark, Denmark.

[8] Shejire DNA project, Abai ave. 150/230, 050046 Almaty, Kazakhstan.

[9] Institute of Archaeology and Steppe Civilization, Al-Farabi Kazakh National University, Almaty, 050040, Kazakhstan.

[10] Margulan Joint Research Center for Archeological Studies, Toraighyrov Pavlodar State. University, Pavlodar, Kazakhstan.

[11] Department of Human Genetics, University of Chicago. Chicago, IL.

[12] Saryarkinsky Institute of Archaeology, Buketov Karaganda State University, Karaganda. 100074, Kazakhstan.

[13] Department of Anthropology, University of Alaska, Fairbanks, USA.

[14] The Institute of Forensic Sciences, Istanbul University, Istanbul, Turkey.

[15] Department of Genetics, Hazara University, Garden Campus, Mansehra, Pakistan.

[16] Department of Historical Studies, University of Gothenburg, 40530 Göteborg, Sweden.

[17] Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, Russia.

[18] Institute for the History of the Material Culture, Russian Academy of Sciences.

[19]Japanese Institute of Anatolian Archaeology, Kaman, Kırşehir, Turkey.

[20]Department of Archaeology, Faculty of Arts, Gazi University, Ankara, Turkey.

[21]Islamia University, Peshawar, Pakistan.

[22]Department of Archaeology, University of Exeter, Exeter, EX4 4QE, UK.

[23]Department of Archeology, Hazara University, Garden Campus, Mansehra, Pakistan.

[24]Directorate of Archaeology and Museums Government of Khyber Pakhtunkhwa, Pakistan.

[25]Archaeological Museum Harappa at Archaeology Department Govt. of Punjab, Pakistan.

[26]Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, Academician Lavrent'iev Ave. 17, Novosibirsk, 630090, Russia.

[27]Department of History, Irkutsk State University, Karl Marx Street 1, Irkutsk 664003, Russia.

[28]Department of Ecology and Evolution,University of Chicago. Chicago, IL.

[29]Department of Anthropology, University of Alberta, Edmonton, Alberta, T6G 2H4, Canada.

[30]Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université deToulouse, Université Paul Sabatier, 31000 Toulouse, France.

[31]Departments of Integrative Biology and Statistics, University of Berkeley, USA.

[32]Department of Zoology, University of Cambridge, UK.

*Correspondence to: rd109@cam.ac.uk (R.D.); ewillerslev@snm.ku.dk (E.W.).

†These authors contributed equally to this work.

Abstract: The Yamnaya expansions from the western steppe into Europe and Asia during the Early Bronze Age (~3000 BCE) are believed to have brought with them Indo-European languages and possibly horse husbandry. We analyze 74 ancient whole-genome sequences from across Inner Asia and Anatolia and show that the Botai people associated with the earliest horse husbandry derived from a hunter-gatherer population deeply diverged from the Yamnaya. Our results also suggest distinct migrations bringing West Eurasian ancestry into South Asia before and after but not at the time of Yamnaya culture. We find no evidence of steppe ancestry in Bronze Age Anatolia from when Indo-European languages are attested there. Thus, in contrast to Europe, Early Bronze Age Yamnaya-related migrations had limited direct genetic impact in Asia.

One Sentence Summary: We investigate the origins of Indo-European languages in Asia by coupling ancient genomics to archaeology and linguistics.

Main Text:

The vast grasslands making up the Eurasian steppe zones, from Ukraine through Kazakhstan to Mongolia, have served as a crossroad for human population movements during the last 5000 years (*1*–*3*), but the dynamics of its human occupation—especially of the earliest period—remain poorly understood. The domestication of the horse at the transition from the Copper Age to the Bronze Age ~3000 BCE, enhanced human mobility (*4*, *5*) and may have triggered waves of migration. According to the "Steppe Hypothesis," this expansion of groups in the western steppe related to the Yamnaya and Afanasievo cultures was associated with the spread of Indo-European (IE) languages into Europe and Asia (*1*, *2*, *4*, *6*). The peoples who formed the Yamnaya and Afanasievo cultures belonged to the same genetically homogenous population, with direct ancestry attributed to both Copper Age (CA) western steppe pastoralists, descending primarily from the European Eastern hunter-gatherers (EHG) of the Mesolithic, and to Caucasian groups (*1*, *2*), related to Caucasus hunter-gatherers (CHG) (*7*).

Within Europe, the "Steppe Hypothesis" is supported by the reconstruction of Proto-IE (PIE) vocabulary (*8*), as well as by archaeological and genomic evidence of human mobility and Early Bronze Age (3000–2500 BCE) cultural dynamics (*9*). For Asia, however, several conflicting interpretations have long been debated. These concern the origins and genetic composition of the local Asian populations encountered by the Yamnaya- and Afanasievo-related populations, including the groups associated with Botai, a site that offers the earliest evidence for horse husbandry (*10*). In contrast, the more western sites that have been supposed by some to reflect the use of horses in the Copper Age (*4*) lack direct evidence of domesticated horses. Even the later use of horses among Yamnaya pastoralists has been questioned by some (*11*) despite the key role of horses in the "Steppe Hypothesis." Furthermore, genetic, archaeological, and linguistic hypotheses diverge on the timing and processes by which steppe genetic ancestry and the IE languages spread into South Asia (*4*, *6*, *12*). Similarly, in present-day Turkey, the emergence of the Anatolian IE language branch including the Hittite language remains enigmatic, with conflicting hypotheses about population migrations leading to its emergence in Anatolia (*4*, *13*).

Ancient genomes inform upon human movements within Asia

We analyzed whole genome sequence data of 74 ancient humans (*14*, *15*) (Tables S1 to S3) ranging from the Mesolithic (~9000 BCE) to Medieval times, spanning ~5000 km across Eastern Europe, Central Asia, and Western Asia (Anatolia) (Fig. 1). Our genome data includes 3 Copper Age individuals (~3500–3300 BCE) from Botai in northern Kazakhstan (Botai_CA; 13.6X, 3.7X, and 3X coverage, respectively), 1 Early Bronze Age (~2900 BCE) Yamnaya sample from Karagash, Kazakhstan*(16)* (YamnayaKaragash_EBA; 25.2X), 1 Mesolithic (~9000 BCE) EHG from Sidelkino, Russia (SidelkinoEHG_ML; 2.9X), 2 Early/Middle Bronze Age (~2200 BCE) central steppe individuals (~4200 BP) (CentralSteppe_EMBA; 4.5X and 9.1X average coverage, respectively) from burials at Sholpan and Gregorievka that display cultural similarities to Yamnaya and Afanasievo (*12*), 19 individuals of the Bronze Age (~2500–2000 BCE) Okunevo culture of the Minusinsk Basin in the Altai region (Okunevo_EMBA; ~1X average coverage; 0.1–4.6X), 31 Baikal Hunter-Gatherer genomes (~1X average coverage; 0.2–4.5X) from the cis-Baikal region bordering on Mongolia and ranging in time from the Early Neolithic (~5200–4200 BCE; Baikal_EN) to the Early Bronze Age (~2200–1800 BCE; Baikal_EBA), 4 Copper Age individuals (~3300–3200 BCE; Namazga_CA; ~1X average coverage; 0.1–2.2X) from Kara-Depe and Geoksur in the Kopet Dag piedmont strip of Turkmenistan, affiliated with the period III cultural layers at Namazga-Depe (Fig. S1), plus 1 Iron Age individual (Turkmenistan_IA; 2.5X) from Takhirbai in the same area dated to ~800 BCE, and 12 individuals from Central Turkey (Figs. S2 to S4), spanning from the Early Bronze Age (~2200 BCE; Anatolia_EBA) to the Iron Age (~600 BCE; Anatolia_IA), and including 5 individuals from presumed Hittite-speaking settlements (~1600 BCE; Anatolia_MLBA), and 2 individuals dated to the Ottoman Empire (1500 CE; Anatolia_Ottoman; 0.3–0.9X). All the population labels including those referring to previously published ancient samples are listed in Table S4 for contextualization. Additionally, we sequenced 41 high-coverage (30X) present-day Central Asian genomes, representing 17 self-declared ethnicities (Fig. S5) as well as collected and SNP-typed 140 individuals from 5 IE-speaking populations in northern Pakistan.

Tests indicated that the contamination proportion of the data was negligible (*14*) (see Table S1), and we removed related individuals from frequency-based statistics (Fig. S6; Table S5). Our high-coverage Yamnaya genome from Karagash is consistent with previously published Yamnaya and Afanasievo genomes, and our Sidelkino genome is consistent with previously published EHG genomes, on the basis that there is no statistically significant deviation from 0 of D-statistics of the form D(Test, Mbuti; SidelkinoEHG_ML, EHG) (Fig. S7) or of the form D(Test, Mbuti; YamnayaKaragash_EBA, Yamnaya) (Fig. S8; additional D-Statistics shown on Figs. S9 to S12).

Genetic origins of local Inner Asian populations

In the Early Bronze Age around 3000 BCE, the Afanasievo culture was formed in the Altai region by people related to the Yamnaya, who migrated 3000 km across the central steppe from the western steppe (*1*), and are often identified as the ancestors of the IE-speaking Tocharians of 1st millennium northwestern China (*4, 6*). At this time, the region they passed through was populated by horse hunter-herders (*4, 10, 17*), while further east the Baikal region hosted groups that had remained hunter-gatherers since the Paleolithic (*18–22*). Subsequently, the Okunevo culture replaced the Afanasievo culture. The genetic origins and relationships of these peoples have been largely unknown (*23, 24*).

To address these issues we characterized the genomic ancestry of the local Inner Asian populations around the time of the Yamnaya and Afanasievo expansion. Comparing our ancient samples to a range of present-day and ancient samples with principal components analysis (PCA), we find that the Botai_CA, CentralSteppe_EMBA, Okunevo_EMBA, and Baikal populations (Baikal_EN and Baikal_EBA) are distributed along a previously undescribed genetic cline. This cline extends from the EHG of the western steppe to the Bronze Age (~2000–1800 BCE) and Neolithic (~5200–4200 BCE) hunter-gatherers of Lake Baikal in Central Asia, which are located on the PCA plot close to modern East Asians and two Early Neolithic (~5700 BCE) Devil's Gate samples (*25*) (Fig. 2, and Fig. S13). In accordance with their position along the west-to-east gradient in the PCA, increased East Asian ancestry is evident in ADMIXTURE model-based clustering (Fig. 3; Figs. S14 and S15) and by D-statistics for Sholpan and Gregorievka (CentralSteppe_EMBA) and Okunevo_EMBA, relative to Botai_CA and the Baikal_EN sample: D(Baikal_EN, Mbuti; Botai_CA, Okunevo_EMBA) = -0.025 Z = -12; D(Baikal_EN, Mbuti; Botai_CA, Sholpan) = -0.028 Z = -8.34; D(Baikal_EN, Mbuti; Botai_CA, Gregorievka) = -0.026 Z = -7.1. The position of this cline suggests that the central steppe Bronze Age populations all form a continuation of the "Ancient North Eurasian" (ANE) population, previously known from the 24-kyr-old Mal'ta (MA1), the 17-kyr-old AG-2 (*26*), and the ~14.7-kyr-old AG-3 (*27*) individuals from Siberia.

To investigate ancestral relationships between these populations, we used coalescent modelling with the momi program (*28*) (Fig. 4; Figs S16 to S22; Tables S6 to S11). This exploits the full joint-site frequency spectrum and can separate genetic drift into divergence-time and population-size components, in comparison to PCA, admixture, and qpAdm approaches, which are based on pairwise covariances. We find that Botai_CA, CentralSteppe_EMBA, Okunevo_EMBA, and Baikal populations are deeply separated from other ancient and present-day populations and are best modelled as mixtures in different proportions of ANE ancestry and an Ancient East Asian (AEA) ancestry component represented by Baikal_EN with mixing times dated to approximately 5000 BCE. Although some modern Siberian samples lie under the Baikal samples in Fig. 2A, these are separated out in a more limited PCA, involving just those populations and the ancient samples (Fig. S23). Our momi model infers that the ANE lineage separated approximately 15 kya in the Upper Paleolithic from the EHG lineage to the west, with no independent drift assigned to MA1. This suggests that MA1 may represent their common ancestor. Similarly, the AEA lineage to the east also separated around 15 kya, with the component that leads to Baikal_EN and the AEA component of the steppe separating from the lineage leading to present-day East Asian populations represented by Han Chinese (Figs. S19 to S21). The ANE and AEA lineages themselves are estimated as having separated approximately 40 kya, relatively soon after the peopling of Eurasia by modern humans.

Since the ANE MA1 sample comes from the same cis-Baikal region as the AEA-derived Neolithic samples analyzed here, we thus document evidence for a population replacement between the Paleolithic and the Neolithic in this region. Furthermore, we observe a shift in genetic ancestry between the Early Neolithic (Baikal_EN) and the Late Neolithic / Bronze Age hunter-gatherers (Baikal_LNBA) (Fig. 2A), with the Baikal_LNBA cluster showing admixture from an ANE-related source. We estimate the ANE related ancestry in the Baikal_LNBA to be around ~5–11% (qpAdm; Table S12 (*2*)), using MA1 as a source of ANE, Baikal_EN as a source of AEA, and a set of 6 outgroups. However, neither MA1 nor any of the other steppe populations lie in the direction of Baikal_LNBA from Baikal_EN on the PCA plot (Fig. S23). This suggests that the new ANE ancestry in Baikal_LNBA stems from an unsampled source. Given that this source may have harbored East Asian ancestry, the contribution may be larger than 10%.

These serial changes in the Baikal populations are reflected in Y-chromosome lineages (Fig. 5A; Figs. S24 to S27; Tables S13 and S14). MA1 carries the R haplogroup, whereas the majority of Baikal_EN males belong to N lineages, which were widely distributed across Northern Eurasia (*29*), and the Baikal_LNBA males all carry Q haplogroups, as do most of the Okunevo_EMBA as well as some present-day Central Asians and Siberians. Mitochondrial haplogroups show less turnover (Fig. 5B; Table S15), which could either indicate male-mediated admixture or reflect bottlenecks in the male population.

The deep population structure among the local populations in Inner Asia around the Copper Age / Bronze Age transition is in line with distinct origins of central steppe hunter-herders related to Botai of the central steppe and those related to Altaian hunter-gatherers of the eastern steppe (*30*). Furthermore, this population structure, which is best described as part of the “Ancient North Eurasian” metapopulation, persisted within Inner Asia from the Upper Paleolithic to the end of the Early Bronze Age. In the Baikal region the results show that at least two genetic shifts occurred: first, a complete population replacement of the Upper Paleolithic hunter-gatherers belonging to the “Ancient North Eurasians” by Early Neolithic communities of Ancient East Asian ancestry And second, an admixture event between the latter and additional members of the “Ancient North Eurasian” clade, occurring during the 1500-year period that separates the Neolithic from the Early Bronze Age. These genetic shifts complement previously observed severe cultural changes in the Baikal region (*18*–*22*).

Relevance for history of horse domestication

The earliest unambiguous evidence for horse husbandry is from the Copper Age Botai hunter-herder culture of the central steppe in Northern Kazakhstan around 3500–3000 BCE (*5, 10, 23, 31–33*). There was extensive debate over whether Botai horses were hunted or herded (*33*), but more recent studies have evidenced harnessing and milking (*10, 17*), the presence of likely corrals, and genetic domestication selection at the horse TRPM1 coat-color locus (*32*). Whilst horse husbandry has been demonstrated at Botai, it is also now clear from genetic studies this was not the source of modern domestic horse stock (*32*). Some have suggested that the Botai were local hunter-gatherers who learnt horse husbandry from an early eastward spread of western pastoralists, such as the Copper Age herders buried at Khvalynsk (~5150–3950 BCE), closely related to Yamnaya and Afanasievo (*17*). Others have suggested an in-situ transition from the local hunter-gatherer community (*5*).

We therefore examined the genetic relationship between Yamnaya and Botai. First, we note that whereas Yamnaya is best modelled as an approximately equal mix of EHG and Caucasian HG ancestry and that the earlier Khvalynsk samples from the same area also show Caucasian ancestry, the Botai_CA samples show no signs of admixture with a Caucasian source (Fig. S14). Similarly, while the Botai_CA have some Ancient East Asian ancestry, there is no sign of this in Khvalynsk or Yamnaya. Our momi model (Fig. 4) suggests that, although YamnayaKaragash_EBA shared ANE ancestry with Botai_CA from MA1 through EHG, their lineages diverge approximately 15,000 years ago in the Paleolithic. According to a parametric bootstrap, the amount of gene flow between YamnayaKaragash_EBA and Botai_CA inferred using the SFS was not significantly different from 0 (p-value 0.18 using 300 parametric bootstraps under a null model without admixture; Fig. S18). Additionally, the best-fitting SFS model without any recent gene flow fits the ratio of ABBA-BABA counts for (SidelkinoEHG_ML, YamnayaKaragash_EBA; Botai_CA, AncestralAllele), with Z-score = 0.45 using a block jackknife for this statistic. Consistent with this, a simple qpGraph model without direct gene flow between Botai_CA and Yamnaya, but with shared EHG-related ancestry between them, fits all f4 statistics (Fig. S28), and qpAdm (*2*) successfully fits models for Yamnaya ancestry without any Botai_CA contribution (Table S12).

The separation between Botai and Yamnaya is further reinforced by a lack of overlap in Y-chromosomal lineages (Fig. 5A). While our YamnayaKaragash_EBA sample carries the R1b1a2a2c1 lineage seen in other Yamnaya and present-day Eastern Europeans, one of the two Botai_CA males belongs to the basal N lineage, whose subclades have a predominantly Northern Eurasian distribution, while the second carries the R1b1a1 haplogroup, restricted almost exclusively to Central Asian and Siberian populations (*34*). Neither of these Botai lineages has been observed among Yamnaya males (Table S13; Fig. S25).

Using chromopainter (*35*) (Figs. S29 to S32) and rare variant sharing (*36*) (Figs. S33 to S35), we also identify a disparity in affinities with present-day populations between our high-coverage Yamnaya and Botai genomes. Consistent with previous results (*1, 2*), we observe a contribution from YamnayaKaragash_EBA to present-day Europeans. Conversely, Botai_CA shows greater affinity to Central Asian, Siberian, and Native American populations, coupled with some sharing with northeastern European groups at a lower level than that for Yamnaya, due to their ANE ancestry.

Further towards the Altai, the genomes of two CentralSteppe_EMBA women, who were buried in Afanasievo-like pit graves, revealed them to be representatives of an unadmixed Inner Asian ANE-related group, almost indistinguishable from the Okunevo_EMBA of the Minusinsk

Basin north of the Altai through D-statistics (Fig. S11). This lack of genetic and cultural congruence may be relevant to the interpretation of Afanasievo-type graves elsewhere in Central Asia and Mongolia (*37*). However, in contrast to the lack of identifiable admixture from Yamnaya and Afanasievo in the CentralSteppe_EMBA, there is an admixture signal of 10–20% Yamnaya and Afanasievo in the Okunevo_EMBA samples (Fig. S21), consistent with evidence of western steppe influence. This signal is not seen on the X chromosome (qpAdm p-value for admixture on X 0.33 compared to 0.02 for autosomes), suggesting a male-derived admixture, also consistent with the fact that 1 of 10 Okunevo_EMBA males carries a R1b1a2a2 Y chromosome related to those found in western pastoralists (Fig. 5). In contrast, there is no evidence of western steppe admixture among the more eastern Baikal region Bronze Age (~2200–1800 BCE) samples (Fig. S14).

The lack of evidence of admixture between Botai horse herders and western steppe pastoralists is consistent with these latter migrating through the central steppe but not settling until they reached the Altai to the east (*4*). More significantly, this lack of admixture suggests that horses were domesticated by hunter-gatherers not previously familiar with farming, as were the cases for dogs (*38*) and reindeer (*39*). Domestication of the horse thus may best parallel that of the reindeer, a food animal that can be milked and ridden, which has been proposed to be domesticated by hunters via the “prey path” (*40*); indeed anthropologists note similarities in cosmological beliefs between hunters and reindeer herders (*41*). In contrast, most animal domestications were achieved by settled agriculturalists (*5*).

Origins of Western Eurasian genetic signatures in South Asians

The presence of Western Eurasian ancestry in many present-day South Asian populations south of the central steppe has been used to argue for gene flow from Early Bronze Age (~3000–2500 BCE) western steppe pastoralists into the region (*42, 43*). However, direct influence of Yamnaya or related cultures of that period is not visible in the archaeological record, except perhaps for a single burial mound in Sarazm in present-day Tajikistan of contested age (*44, 45*). Additionally, linguistic reconstruction of proto-culture coupled with the archaeological chronology evidences a Late (~2300–1200 BCE) rather than Early Bronze Age (~3000–2500 BCE) arrival of the Indo-Iranian languages into South Asia (*16, 45, 46*). Thus, debate persists as to how and when Western Eurasian genetic signatures and IE languages reached South Asia.

To address these issues, we investigated whether the source of the Western Eurasian signal in South Asians could derive from sources other than Yamnaya and Afanasievo (Fig. 1). Both Early Bronze Age (~3000–2500 BCE) steppe pastoralists Yamnaya and Afanasievo and Late Bronze Age (~2300–1200 BCE) Sintashta and Andronovo carry substantial amounts of EHG and CHG ancestry (*1, 2, 7*), but the latter group can be distinguished by a genetic component acquired through admixture with European Neolithic farmers during the formation of the Corded Ware complex (*1, 2*), reflecting a secondary push from Europe to the east through the forest-steppe zone.

We characterized a set of 4 south Turkmenistan samples from Namazga period III (~3300 BCE). In our PCA analysis, the Namazga_CA individuals were placed in an intermediate position between Iran Neolithic and Western Steppe clusters (Fig. 2). Consistent with this, we find that the Namazga_CA individuals carry a significantly larger fraction of EHG-related ancestry than Neolithic skeletal material from Iran (D(EHG, Mbuti; Namazga_CA, Iran_N) Z =

4.49), and we are not able to reject a two-population qpAdm model in which Namazga_CA ancestry was derived from a mixture of Neolithic Iranians and EHG (~21%; p = 0.49).

Although CHG contributed both to Copper Age steppe individuals (e.g., Khvalynsk ~5150–3950 BCE) and substantially to Early Bronze Age (~3000–2500 BCE) steppe Yamnaya and Afanasievo (*1, 2, 7, 47*), we do not find evidence of CHG-specific ancestry in Namazga. Despite the adjacent placement of CHG and Namazga_CA on the PCA plot, D(CHG, Mbuti; Namazga_CA, Iran_N) does not deviate significantly from 0 (Z = 1.65), in agreement with ADMIXTURE results (Fig. 3; Fig. S14). Moreover, a three-population qpAdm model using Iran Neolithic, EHG, and CHG as sources yields a negative admixture coefficient for CHG. This suggests that while we cannot totally reject a minor presence of CHG ancestry, steppe-related admixture most likely arrived in the Namazga population prior to the Copper Age or from unadmixed sources related to EHG. This is consistent with the upper temporal boundary provided by the date of the Namazga_CA samples (~3300 BCE). In contrast, the Iron Age (~900–200 BCE) individual from the same region as Namazga (sample DA382, labelled Turkmenistan_IA) is closer to the steppe cluster in the PCA plot and does have CHG-specific ancestry. However, it also has European farmer-related ancestry typical of Late Bronze Age (~2300–1200 BCE) steppe populations (*1–3, 47*) (D(Neolithic European, Mbuti; Namazga_CA, Turkmenistan_IA) Z = -4.04), suggesting that it received admixture from Late (~2300–1200 BCE) rather than Early Bronze Age (~3000–2500 BCE) steppe populations.

In a PCA focused on South Asia (Fig. 2B), the first dimension corresponds approximately to West-East and the second dimension to North-South. Near the lower right are the Andamanese Onge previously used to represent the “Ancient South Asian” component (*12, 42*). Contemporary South Asian populations are placed along both East-West and North-South gradients, reflecting the presence of three major ancestry components in South Asia deriving from “West Eurasians,” “South Asians,” and “East Asians.” Since the Namazga_CA individuals appear at one end of the West Eurasian / South Asian axis, and given their geographical proximity to South Asia, we tested this group as a potential source in a set of qpAdm models for the South Asian populations (Fig. 6).

We are not able to reject a two-population qpAdm model using Namazga_CA and Onge for 9 modern southern and predominantly Dravidian-speaking populations (Fig. 6; Fig. S36; Tables S16 and S17). In contrast, for 7 other populations belonging to the northernmost Indic- and Iranian-speaking groups this two-population model is rejected, but not a three-population model including an additional Late Bronze Age (~2300–1200 BCE) steppe source. Lastly, for 7 southeastern Asian populations, 6 of which were Tibeto-Burman or Austro-Asiatic speakers, the three-population model with Late Bronze Age (~2300–1200 BCE) steppe ancestry was rejected, but not a model in which Late Bronze Age (~2300–1200 BCE) steppe ancestry was replaced with an East Asian ancestry source, as represented by the Late Iron Age (~200 BCE–100 CE) Xiongnu (Xiongnu_IA) nomads from Mongolia (*3*). Interestingly, for two northern groups, the only tested model we could not reject included the Iron Age (~900–200 BCE) individual (Turkmenistan_IA) from the Zarafshan Mountains and the Xiongnu_IA as sources. These findings are consistent with the positions of the populations in PCA space (Fig. 2B), and further supported by ADMIXTURE analysis (Fig. 3) with two minor exceptions: in both the Iyer and the Pakistani Gujar we observe a minor presence of the Late Bronze Age (~2300–1200 BCE) steppe ancestry component (Fig. S14) not detected by the qpAdm approach. Additionally, we document admixture along the “West Eurasian” and “East Asian” clines of all South Asian populations using D-statistics (Fig. S37).

Thus, we find that ancestries deriving from 4 major separate sources fully reconcile the population history of present-day South Asians (Figs. 3 and 6), one anciently South Asian, one from Namazga or a related population, a third from Late Bronze Age (~2300–1200 BCE) steppe pastoralists, and lastly one from East Asia. They account for western ancestry in some Dravidian populations that lack CHG-specific ancestry while also fitting the observation that whenever there is CHG-specific ancestry and considerable EHG ancestry there is also European Neolithic ancestry (Fig. 3). This implicates Late Bronze Age (~2300–1200 BCE) steppe rather than Early Bronze Age (~3000–2500 BCE) Yamnaya and Afanasievo admixture into South Asia. The proposal that the IE steppe ancestry arrived in the Late Bronze Age (~2300–1200 BCE) is also more consistent with archaeological and linguistic chronology (*44*, *45*, *48*, *49*). Thus, it seems that the Yamnaya- and Afanasievo-related migrations did not have a direct genetic impact in South Asia.

Lack of steppe genetic impact in Anatolians

Finally, we consider the evidence for Bronze Age steppe genetic contributions in West Asia. There are conflicting models for the earliest dispersal of IE languages into Anatolia (*4*, *50*). The now extinct Bronze Age Anatolian language group represents the earliest historically attested branch of the IE language family and is linguistically held to be the first branch to have split off from PIE (*53*, *54*, *58*). One key question is whether Proto-Anatolian is a direct linguistic descendant of the hypothesized Yamnaya PIE language or whether Proto-Anatolian and the PIE language spoken by Yamnaya were branches of a more ancient language ancestral to both (*49*, *53*). Another key question relates to whether Proto-Anatolian speakers entered Anatolia as a result of a “Copper Age western steppe migration” (~5000–3000 BCE) involving movement of groups through the Balkans into Northwest Anatolia (*4*, *71*, *73*), or a “Caucasian” route that links language dispersal to intensified north-south population contacts facilitated by the trans-Caucasian Maykop culture around 3700–3000 BCE (*50*, *54*).

Ancient DNA findings suggest extensive population contact between the Caucasus and the steppe during the Copper Age (~5000–3000 BCE) (*1*, *2*, *42*). Particularly, the first identified presence of Caucasian genomic ancestry in steppe populations is through the Khvalynsk burials (*2*, *47*) and that of steppe ancestry in the Caucasus is through Armenian Copper Age individuals (*42*). These admixture processes likely gave rise to the ancestry that later became typical of the Yamnaya pastoralists (*7*), whose IE language may have evolved under the influence of a Caucasian language, possibly from the Maykop culture (*50*, *55*). This scenario is consistent with both the “Copper Age steppe” (*4*) and the “Caucasian” models for the origin of the Proto-Anatolian language (*56*).

The PCA (Fig. 2B) indicates that all the Anatolian genome sequences from the Early Bronze Age (~2200 BCE) and Late Bronze Age (~1600 BCE) cluster with a previously sequenced Copper Age (~3900–3700 BCE) individual from Northwestern Anatolia and lie between Anatolian Neolithic (Anatolia_N) samples and CHG samples but not between Anatolia_N and EHG samples. A test of the form D(CHG, Mbuti; Anatolia_EBA, Anatolia_N) shows that these individuals share more alleles with CHG than Neolithic Anatolians do (Z = 3.95), and we are not able to reject a two-population qpAdm model in which these groups derive ~60% of their ancestry from Anatolian farmers and ~40% from CHG-related ancestry (p-value =

0.5). This signal is not driven by Neolithic Iranian ancestry, since the result of a similar test of the form D(Iran_N, Mbuti; Anatolia_EBA, Anatolia_N) does not deviate from zero (Z = 1.02). Taken together with recent findings of CHG ancestry on Crete (*57*), our results support a widespread CHG-related gene flow, not only into Central Anatolia but also into the areas surrounding the Black Sea and Crete. The latter are not believed to have been influenced by steppe-related migrations and may thus correspond to a shared archaeological horizon of trade and innovation in metallurgy (*66*).

Importantly, a test of the form D(EHG, Mbuti; Anatolia_EBA, Anatolia_MLBA) supports that the Central Anatolian gene pools, including those sampled from settlements thought to have been inhabited by Hittite speakers, were not impacted by steppe populations during the Early and Middle Bronze Age (Z = -1.83). Both of these findings are further confirmed by results from clustering analysis (Fig. 3). The CHG-specific ancestry and the absence of EHG-related ancestry in Bronze Age Anatolia would be in accordance with intense cultural interactions between populations in the Caucasus and Anatolia observed during the late 5th millennium BCE that seem to come to an end in the first half of the 4th millennium BCE with the village-based egalitarian Kura-Araxes' society (*59, 60*), thus preceding the emergence and dispersal of Proto-Anatolian.

Our results indicate that the early spread of IE languages into Anatolia was not associated with any large-scale steppe-related migration, as previously suggested (*61*). Additionally, and in agreement with the later historical record of the region (*62*), we find no correlation between genetic ancestry and exclusive ethnic or political identities among the populations of Bronze Age Central Anatolia, as has previously been hypothesized (*63*).

Discussion

For Europe, ancient genomics have revealed extensive population migrations, replacements, and admixtures from the Upper Paleolithic to the Bronze Age (*1, 2, 27, 64, 65*), with a strong influence across the continent from the Early Bronze Age (~3000–2500 BCE) western steppe Yamnaya. In contrast, for Central Asia, continuity is observed from the Upper Paleolithic to the end of the Copper Age (~3500–3000 BCE), with descendants of Paleolithic hunter-gatherers persisting as largely isolated populations after the Yamnaya and Afanasievo pastoralist migrations. Instead of western pastoralists admixing with or replacing local groups, we see groups with East Asian ancestry replacing ANE populations in the Lake Baikal region. Thus, unlike in Europe, the hunter/gathering/herding groups of Inner Asia were much less impacted by the Yamnaya and Afanasievo expansion. This may be due to the rise of early horse husbandry, likely initially originated through a local "prey route" (*40*) adaptation by horse-dependent hunter-gatherers at Botai. Since work on ancient horse genomes (*32*) indicates that Botai horses were not the main source of modern domesticates, this suggests the existence of a second center of domestication, but whether this second center was associated with the Yamnaya and Afanasievo cultures remains uncertain in the absence of horse genetic data from their sites.

Our finding that the Copper Age (~3300 BCE) Namazga-related population from the borderlands between Central and South Asia contains both "Iran Neolithic" and EHG ancestry but not CHG-specific ancestry provides a solution to problems concerning the Western Eurasian genetic contribution to South Asians. Rather than invoking varying degrees of relative contribution of "Iran Neolithic" and Yamnaya ancestries, we explain the two western genetic components with two separate admixture events. The first event, potentially prior to the Bronze

Age, spread from a non-IE-speaking farming population from the Namazga culture or a related source down to Southern India. Then the second came during the Late Bronze Age (~2300–1200 BCE) through established contacts between pastoral steppe nomads and the Indus Valley, bringing European Neolithic as well as CHG-specific ancestry, and with them Indo-Iranian languages into northern South Asia. This is consistent with a long-range South Eurasian trade network around 2000 BCE (*4*), shared mythologies with steppe-influenced cultures (*41*, *60*), linguistic relationships between Indic spoken in South Asia, and written records from Western Asia from the first half of the 18th century BCE onwards (*49, 52*).

In Anatolia, our samples do not genetically distinguish Hittite and other Bronze Age Anatolians from an earlier Copper Age sample (~3943-3708 BCE). All these samples contain a similar level of CHG ancestry but no EHG ancestry. This is consistent with Anatolian / Early European farmer ancestry, but not steppe ancestry, in the Copper Age Balkans (*67*) and implies that the Anatolian clade of IE languages did not derive from a large-scale Copper Age / Early Bronze Age population movement from the steppe (contra (*4*)). Our findings are thus consistent with historical models of cultural hybridity and “Middle Ground” in a multi-cultural and multi-lingual but genetically homogenous Bronze Age Anatolia (*68, 69*).

Current linguistic estimations converge on dating the Proto-Anatolian split from residual PIE to the late 5th or early 4th millennia BCE (*58, 70*) and place the breakup of Anatolian IE inside Turkey prior to the mid-3rd millennium (*53, 71, 72*). In (*49*) we present new onomastic material (*51*) that pushes the period of Proto-Anatolian linguistic unity even further back in time. We cannot at this point reject a scenario in which the introduction of the Anatolian IE languages into Anatolia was coupled with the CHG-derived admixture prior to 3700 BCE, but note that this is contrary to the standard view that PIE arose in the steppe north of the Caucasus (*4*) and that CHG ancestry is also associated with several non-IE-speaking groups, historical and current. Indeed, our data are also consistent with the first speakers of Anatolian IE coming to the region by way of commercial contacts and small-scale movement during the Bronze Age. Among comparative linguists, a Balkan route for the introduction of Anatolian IE is generally considered more likely than a passage through the Caucasus, due, for example, to greater Anatolian IE presence and language diversity in the west (*73*). Further discussion of these options is given in the archaeological and linguistic supplementary discussions (*48, 49*).

Thus, while the “Steppe hypothesis,” in the light of ancient genomics, has so far successfully explained the origin and dispersal of IE languages and culture in Europe, we find that several elements must be re-interpreted to account for Asia. First, we show that the earliest unambiguous example of horse herding emerged amongst hunter-gatherers, who had no significant genetic interaction with western steppe herders. Second, we demonstrate that the Anatolian IE language branch, including Hittite, did not derive from a substantial steppe migration into Anatolia. And third, we conclude that Early Bronze Age steppe pastoralists did not migrate into South Asia but that genetic evidence fits better with the Indo-Iranian IE languages being brought to the region by descendants of Late Bronze Age steppe pastoralists.

References and Notes:

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).

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).
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* (accepted).
4. D. W. Anthony, *The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world* (Princeton University 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 University 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).
8. J. P. Mallory, D. Q. Adams, *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world* (Oxford University 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).

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).
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).
13. K. Kristiansen, *Europe before History* (Cambridge University Press, Cambridge, 1998).
14. See Supplementary Online 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. *bioRxiv* DOI XXX (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).
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).
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).
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).

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).
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).
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).
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).
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).
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* (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).

35. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).

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).

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).

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).

40. M. A. Zeder, Pathways to animal domestication. *Biodivers. Agric. Domest. Evol Sustain*. 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).

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).

43. 1. 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).

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. Kuzʹmina, *The origin of the Indo-Iranians* (Brill, Leiden, 2007).

46. A. Parpola, *The roots of Hinduism: the early Aryans and the Indus civilization* (Oxford University 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).

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. *bioRxiv* DOI XXX (2018).

49. G. Kroonen, G. Barjamovic, M. Peyrot, Linguistic supplement to Damgaard et al. 2018: Early Indo-European Languages, Anatolian, Tocharian and Indo-Iranian. *bioRxiv* DOI XXX (2018).

50. K. Kristiansen, T. B. Larsson, *The rise of Bronze Age society: travels, transmissions and transformations* (Cambridge University 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 University Press, Stanford, Calif., 1991), pp. 51–66.

55. F. Kortlandt, CC 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. Wash. DC.* **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 University Press, 2015).

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

65. Lazaridis, N. 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).

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. Boroneanţ, 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).

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. 171–88.

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 genet inheritance: problems in comparative linguistics*, A.Y. Aikhenvald, R. M. W. Dixon, Eds. (Oxford University 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 University 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).

77. D. Ferembach, I. Schwindezky, M. Stoukal, Recommendations for age and sex diagnoses of skeletons. *J. Hum. Evol.* **9**, 517–549 (1980).

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).

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).

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 age at death. *Am. J. Phys. Anthropol.* **68**, 15–28 (1985).

81. D. R. Brothwell, *Digging up Bones* (Oxford University 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).

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**, 932–42 (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** (295), 20–30 (2003).

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 [A cemetery at the mouth of the Ida River in the Angara River Valley]. *Arkheologiia Paleoekologiia Ehnologiia Sib Dalnego Vostoka* **1**, 85–90 (1996).

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** (2), 789–799 (2014).

96. T. Nomokonova, R. J. Losey, 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).

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** (3), 991–1008 (2014).

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 University 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).

111. E. Willerslev, A. Cooper, Ancient DNA. *Proc. R. Soc. Lond. B. Biol. Sci.* **272**, 3–16 (2005).

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).

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 (2017).

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).

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).

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).

117. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

118. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997 (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).

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).

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).

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).

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).

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).

125. M. Lipatov, K. Sanjeev, R. Patro, K. Veeramah K. Maximum likelihood estimation of biological relatedness from low coverage sequencing data. *bioRxiv.* 023374 (2015).

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).

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).

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).

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

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

131. P. F. Palamara, L. C. Francioli, P. R. Wilton, G. Genovese, A. Gusev, H. K. Finucane, S. Sankararaman, Genome of the Netherlands Consortium, S. R. Sunyaev, P. I. W. de Bakker, J. Wakeley, I. Pe’er, A. L. Price, Leveraging Distant Relatedness to Quantify Human Mutation and Gene-Conversion Rates. *Am. J. Hum. Genet.* **97**, 775–789 (2015).

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).

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).

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).

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).

136. G. D. Poznik, yHaplo: identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men, *bioRxiv.* 088716 (2016).

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).

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).

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).

140. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32**, 1792–1797 (2004).

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).

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).

144. T. Kivisild, The study of human Y chromosome variation through ancient DNA. *Hum Genet*. **136**, 529–546 (2017).

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).

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).

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).

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).

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).

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**, W58–63 (2016).

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

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).

154. P. Flegontov, N. E. Altinisik, P. Changmai, N. Rohland, Paleo-Eskimo genetic legacy across North America. *bioRxiv.* 203018 (2017).

155. A. T. Duggan, M. Whitten, V. Wiebe, M. Crawford, 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).

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).

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).

158. 1. 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).

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).

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).

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).

162. 1. 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).

163. 1. I. Mathieson, G. McVean, Demography and the age of rare variants. *PLoS Genet.* **10**, e1004528 (2014).

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).

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

167. M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, "Cluster: cluster 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).

Acknowledgments: We thank Kim Magnussen, Lillian Petersen, Cecilie Mortensen, and Andaine Seguin Orlando at the Danish National Sequencing Centre for conducting the sequencing, Paula Reimer and Stephen Hoper at the 14Chrono Center Belfast for providing the AMS dating. We thank Sturla Ellingvåg, Bettina Elisabeth Heyerdahl, and the Explico-Historical Research Foundation team as well as Niobe Thompson for involvement in field work. We thank the Turkish Ministry of Culture and Tourism, Kaman-Kalehöyük Archaeology Museum, and Nevşehir Museum for the permission to samples of Kaman-Kalehöyük and Ovaören. We thank Jesper Stenderup, Pernille V. Olsen, and Tina Brand for technical assistance in the laboratory. We thank Thorfinn Korneliussen for helpful discussions. We thank the St. Johns College in Cambridge for providing the settings for fruitful scientific discussions. We thank all involved archaeologists, historians, and collaborators from Pakistan who assisted IU in the field. We thank Gabit Baimbetov (Shejire DNA), Ilyas Baimukhan, Batyr Daulet, Adbul Kusaev, Ainur Kopbassarova, Youldash Yousupov, Maksum Akchurin, and Vladimir Volkov for important assistance in the field. **Funding:** The study was supported by the Lundbeck Foundation (EW), the Danish National Research Foundation (EW), and KU2016 (EW). Research at the Sanger Institute was supported by the Wellcome Trust (grant 206194). RM was supported by an EMBO Long-Term Fellowship (ALTF 133-2017). JK was supported by the Human Frontiers Science Program (LT000402/2017). Botai fieldwork was supported by University of Exeter, Archaeology Exploration Fund and Niobe Thompson, Clearwater Documentary. AB was supported by NIH grant 5T32GM007197-43. GK was funded by Riksbankens Jubileumsfond and European Research Council. MP was funded by Netherlands Organization for Scientific Research (NWO), project number 276-70-028, IU was funded by the Higher education commission of Pakistan. Archaeological materials from Sholpan and Grigorievka were obtained with partial financial support of the budget program of the Ministry of Education and Science of the Republic of Kazakhstan "Grant financing of scientific research for 2018-2020" No. AP05133498 "Early Bronze Age of the Upper Irtysh". **Author contributions:** EW, KK, AO, and AW: initiated the study. EW, RD, KK, AO, and PBD: designed the study. EW and RD: led the study. KK and AO: led the archaeological part of the study. GK, MP, and GB: led the linguistic part of the study. PBD, CZ, FEY, IU, CdF, MI, HS, ASO, and MEA: produced data. PBD, RM, JK, JVMM, SR, KHI, MS, RN, AB, JN, EW, and RD: analyzed or assisted in analysis of data. PBD, RM, JK, JVMM, RD, EW, AO, KK, GK, MP, GB, BH, MS, and RN: interpreted the results. PBD, EW, RM, RD, AO, GK, JK, GB, JVMM, KK, and MP: wrote the manuscript with considerable input from BH, MS, MEA, and RN. PBD, VZ, VM, IM, NB, EU, VL, FEY, IU, AM, KGS, VM, AG, SO, SYS, CM, HA, AH, AS, NG, MHK, AW, LO, and AO: excavated, curated, sampled and/or described analyzed skeletons. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Genomic data are available for download at the ENA (European Nucleotide Archive) with accession numbers ERP107300 and PRJEB26349. Novel SNP array data from Pakistan can be obtained from EGA through the accession number EGAS00001002965. Y-chromosome and mtDNA data are available at Zenodo under DOI 10.5281/zenodo.1219431.

Supplementary Materials:

Supplementary Text

Figures S1–S37

Tables S1–S17

References (*74*–*168*)

Fig. 1. Geographic location and dates of ancient samples. A) Location of the 74 samples from the steppe, Lake Baikal region, Turkmenistan, and Anatolia analyzed in the present study. MA1, KK1, and Xiongnu_IA were previously published. Geographical background colors indicate the western steppe (pink), central steppe (orange) and eastern steppe (gray). B) Timeline in years before present (BP) for each sample. ML – Mesolithic, EHG – Eastern hunter-gatherer, EN – Early Neolithic, LN – Late Neolithic, CA – Copper Age, EBA – Early Bronze Age, EMBA – Early/Middle Bronze Age, MLBA – Middle/Late Bronze Age, IA – Iron Age.

Fig. 2. Principal component analyses using ancient and present-day genetic data. A) PCA of ancient and modern Eurasian populations. The ancient steppe ancestry cline from EHG to Baikal_EN is visible at the top outside present-day variation, while the YamnayaKaragash_EBA sample has additional CHG ancestry and locates to the left with other Yamnaya and Afanasievo samples. Additionally, a shift in ancestry is observed between the Baikal_EN and Baikal_LNBA, consistent with an increase in ANE-related ancestry in Baikal_LNBA. B) PCA estimated with a subset of Eurasian ancient individuals from the steppe, Iran, and Anatolia as well as present-day South Asian populations. PC1 and PC2 broadly reflect West-East and North-South geography, respectively. Multiple clines of different ancestry are seen in the South Asians, with a prominent cline even within Dravidians in the direction of the Namazga_CA group, which is positioned above Iranian Neolithic in the direction of EHG. In the later Turkmenistan_IA sample, this shift is more pronounced and towards Steppe EBA and MLBA. The Anatolia_CA, EBA and MLBA samples are all between Anatolia Neolithic and CHG, not in the direction of steppe samples.

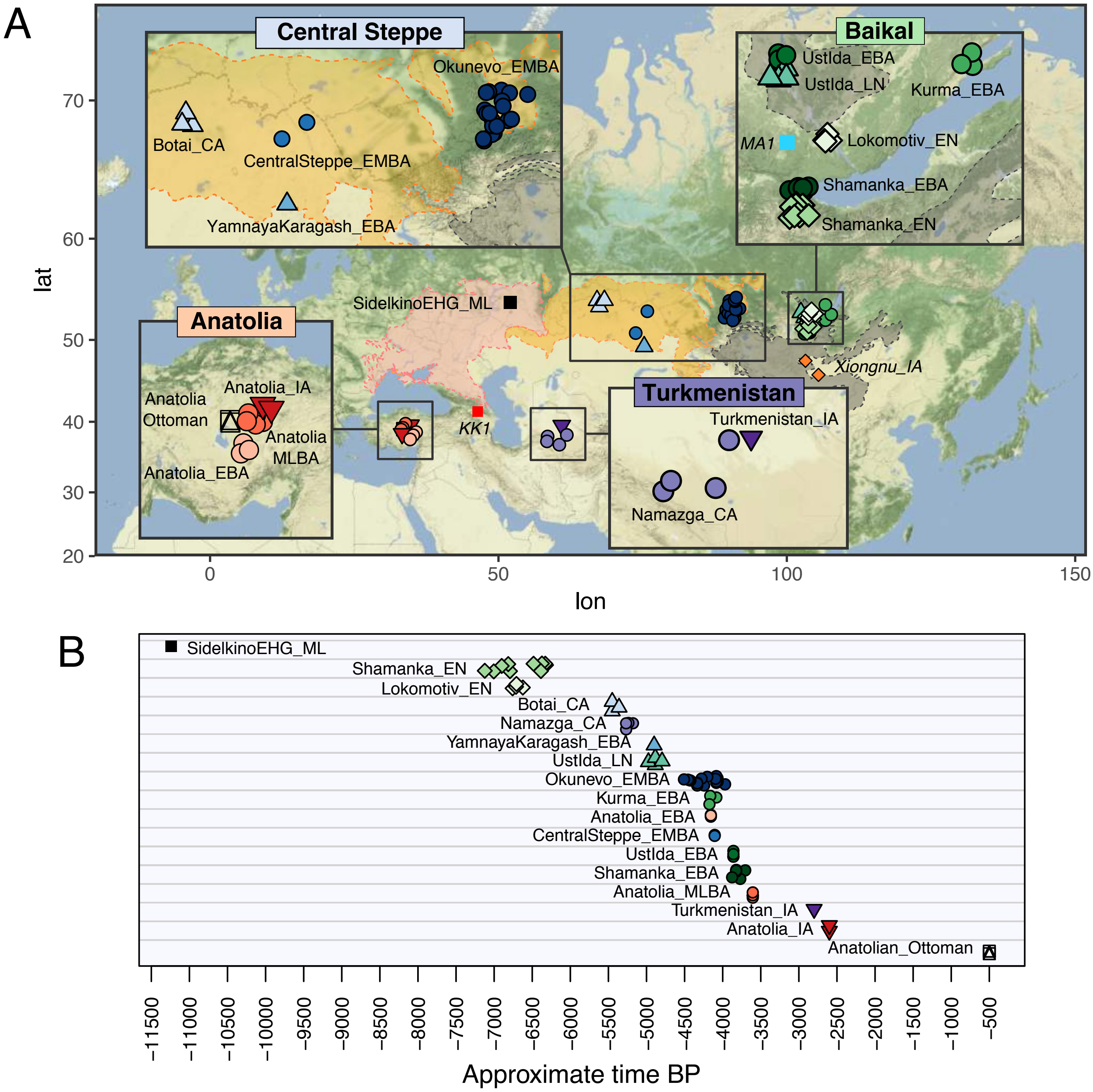
Fig. 3. Model-based clustering analysis of present-day and ancient individuals assuming K = 6 ancestral components. The main ancestry components at K = 6 correlate well with CHG (turquoise), a major component of Iran_N, Namazga_CA and South Asian clines; EHG (pale blue), a component of the steppe cline and present in South Asia; East Asia (yellow ochre), the other component of the steppe cline also in Tibeto-Burman South Asian populations; South Indian (pink), a core component of South Asian populations; Anatolian_N (purple), an important component of Anatolian Bronze Age and Steppe_MLBA; Onge (dark pink) forms its own component.

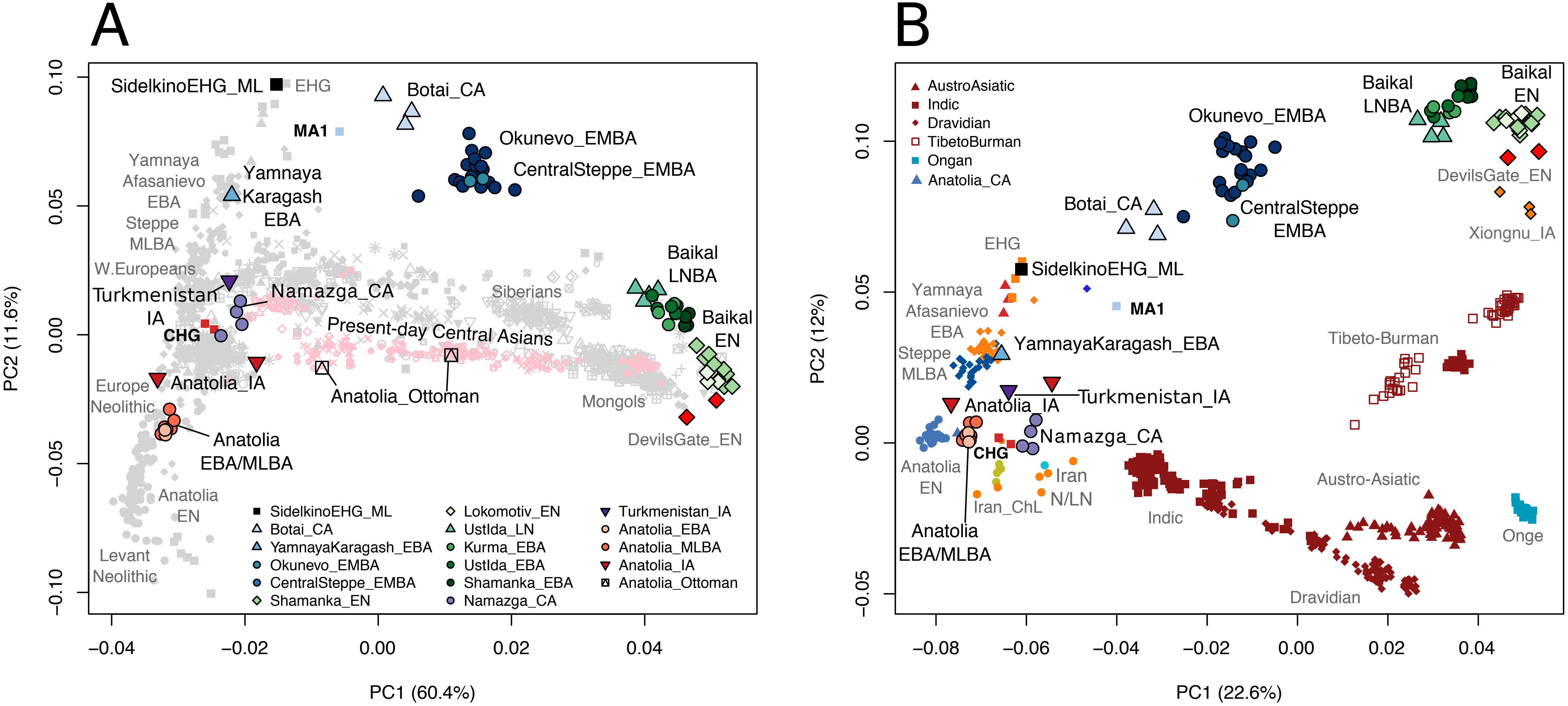
Fig. 4. Demographic model of 10 populations inferred by maximizing the likelihood of the site frequency spectrum (implemented in momi).We used 300 parametric bootstrap simulations (shown in gray transparency) to estimate uncertainty. Bootstrap estimates for the bias and standard deviation of admixture proportions are listed beneath their point estimates. Note that the uncertainty may be underestimated here, due to simplifications or additional uncertainty in the model specification.

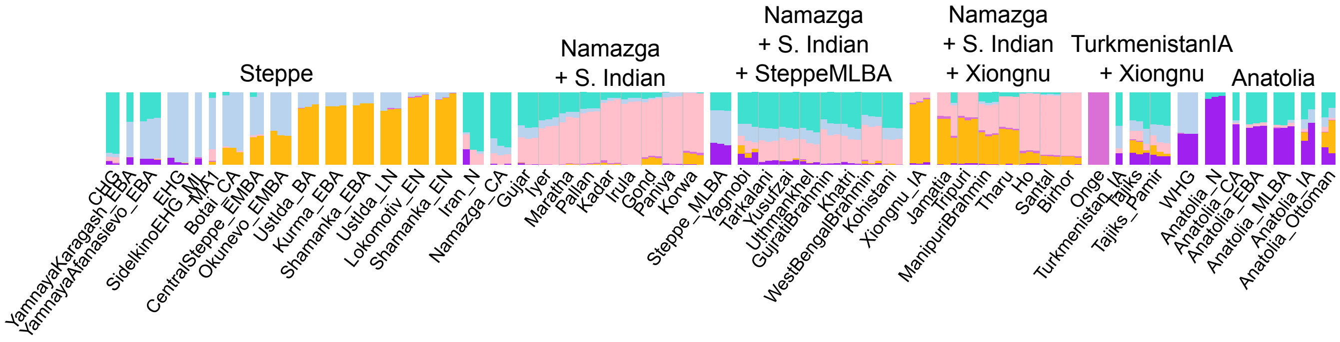
Fig. 5. Y-chromosome and mitochondrial lineages identified in ancient and present-day individuals. A) Maximum likelihood Y-chromosome phylogenetic tree estimated with data from 109 high coverage samples. Dashed lines represent the upper bound for the inclusion of 42 low coverage ancient samples in specific Y-chromosome clades on the basis of the lineages

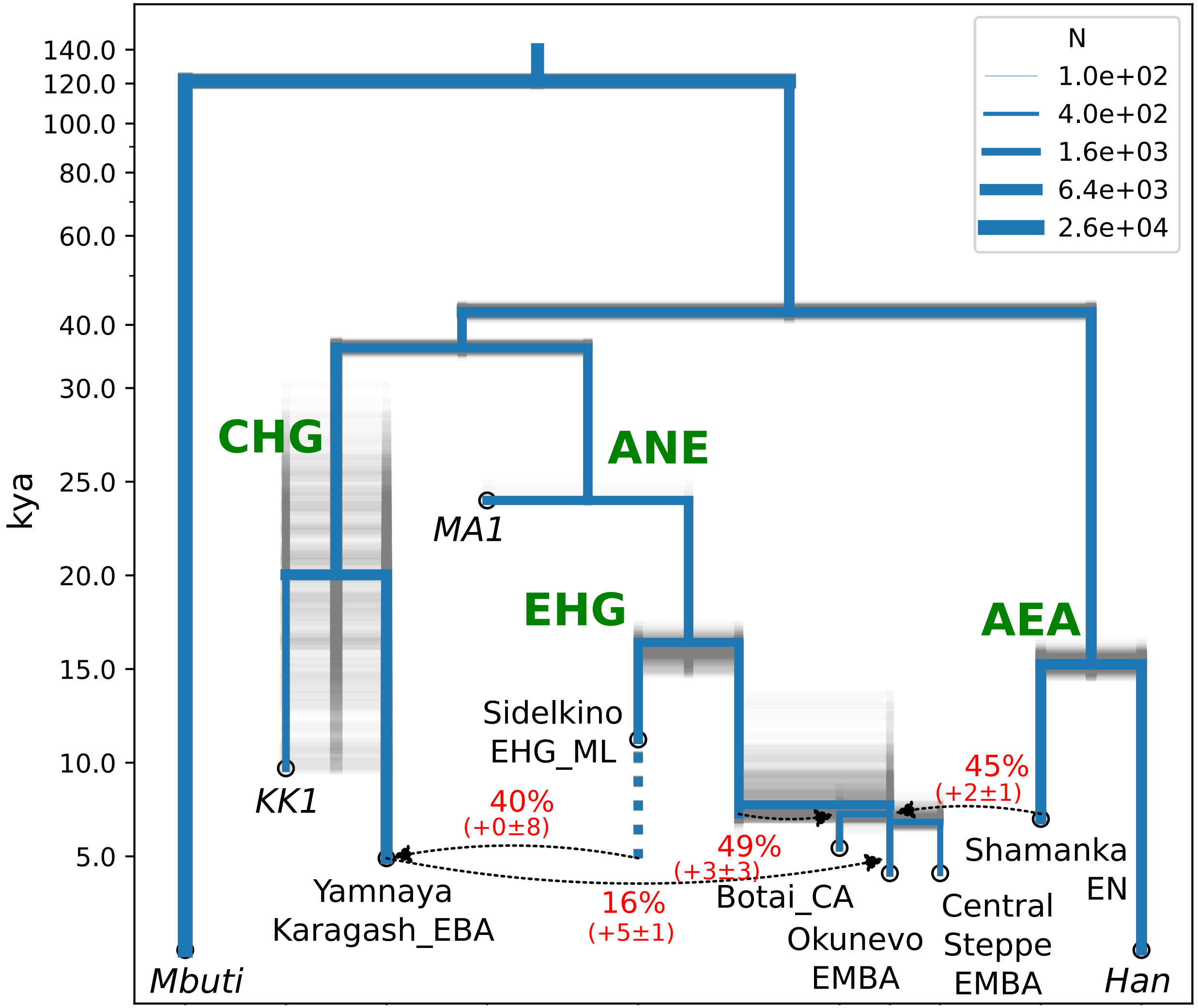
identified. B) Maximum likelihood mitochondrial phylogenetic tree estimated with 182 present-day and ancient individuals. The phylogenies displayed were restricted to a subset of clades relevant to the present work. Columns represent archaeological groups analyzed in the present study, ordered by time, and colored areas indicate membership of the major Y-chromosome and mtDNA haplogroups.

Fig. 6. A summary of the four qpAdm models fitted for South Asian populations. For each modern South Asian population, we fit different models with qpAdm to explain their ancestry composition using ancient groups and present the first model that we could not reject in the following priority order: 1. Namazga_CA + Onge, 2. Namazga_CA + Onge + Late Bronze Age Steppe, 3. Namazga_CA + Onge + Xiongnu_IA (East Asian proxy), and 4. Turkmenistan_IA + Xiongnu_IA. Xiongnu_IA were used here to represent East Asian ancestry. We observe that while South Asian Dravidian speakers can be modelled as a mixture of Onge and Namazga_CA, an additional source related to Late Bronze Age steppe groups is required for IE speakers. In Tibeto-Burman and Austro-Asiatic speakers, an East Asian rather than a Steppe_MLBA source is required.

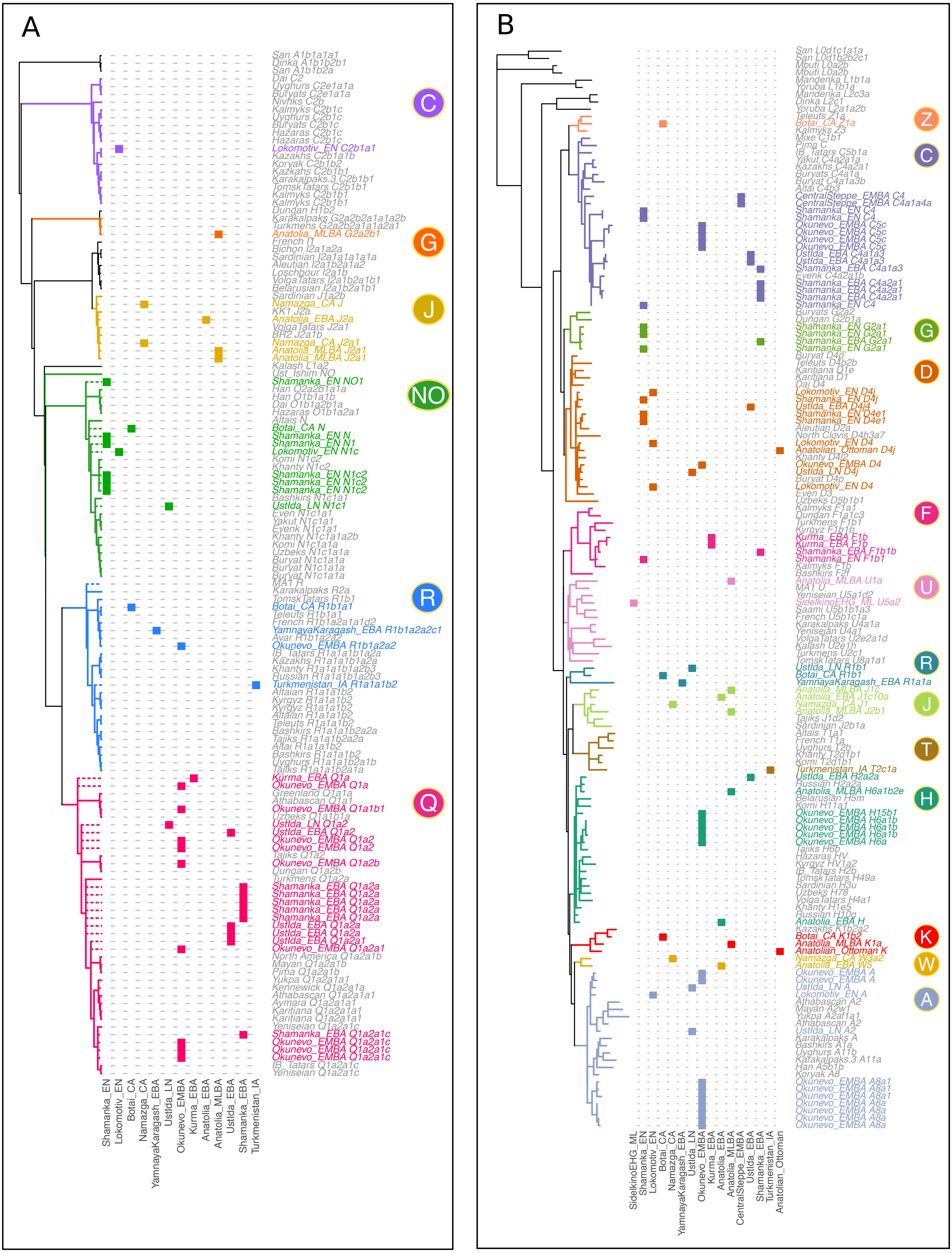

A
Central Steppe
Okunevo_EMBA
Botai_CA
CentralSteppe_EMBA
YamnayaKaragash_EBA
Baikal
UstIda_EBA
UstIda_LN
Kurma_EBA
MA1
Lokomotiv_EN
Shamanka_EBA
Shamanka_EN
SidelkinoEHG_ML
Xiongnu_IA
Anatolia
Anatolia_IA
Anatolia Ottoman
Anatolia MLBA
Anatolia_EBA
KK1
Turkmenistan
Turkmenistan_IA
Namazga_CA
lat
70
60
50
40
30
20
0
50
100
150
lon
B
SidelkinoEHG_ML
Shamanka_EN
Lokomotiv_EN
Botai_CA
Namazga_CA
YamnayaKaragash_EBA
UstIda_LN
Okunevo_EMBA
Kurma_EBA
Anatolia_EBA
CentralSteppe_EMBA
UstIda_EBA
Shamanka_EBA
Anatolia_MLBA
Turkmenistan_IA
Anatolia_IA
Anatolian_Ottoman
−11500
−11000
−10500
−10000
−9500
−9000
−8500
−8000
−7500
−7000
−6500
−6000
−5500
−5000
−4500
−4000
−3500
−3000
−2500
−2000
−1500
−1000
−500
Approximate time BP


A
SidelkinoEHG_ML
EHG
Botai_CA
MA1
Okunevo_EMBA
CentralSteppe_EMBA
Yamnaya Afasanievo EBA
Steppe MLBA
Yamnaya Karagash EBA
W.Europeans
Turkmenistan IA
Namazga_CA
Siberians
Baikal LNBA
Baikal EN
CHG
Present-day Central Asians
Europe Neolithic
Anatolia_IA
Anatolia_Ottoman
Mongols
DevilsGate_EN
Anatolia EBA/MLBA
Anatolia EN
Levant Neolithic
SidelkinoEHG_ML
Botai_CA
YamnayaKaragash_EBA
Okunevo_EMBA
CentralSteppe_EMBA
Shamanka_EN
Lokomotiv_EN
UstIda_LN
Kurma_EBA
UstIda_EBA
Shamanka_EBA
Namazga_CA
Turkmenistan_IA
Anatolia_EBA
Anatolia_MLBA
Anatolia_IA
Anatolia_Ottoman
PC2 (11.6%)
PC1 (60.4%)
0.10
0.05
0.00
−0.05
−0.10
−0.04
−0.02
0.00
0.02
0.04
B
AustroAsiatic
Indic
Dravidian
TibetoBurman
Ongan
Anatolia_CA
Baikal LNBA
Baikal EN
Okunevo_EMBA
DevilsGate_EN
Xiongnu_IA
Botai_CA
CentralSteppe EMBA
EHG
SidelkinoEHG_ML
Yamnaya Afasanievo EBA
MA1
YamnayaKaragash_EBA
Steppe MLBA
Tibeto-Burman
Anatolia_IA
Turkmenistan_IA
Namazga_CA
CHG
Anatolia EN
Iran N/LN
Iran_ChL
Indic
Austro-Asiatic
Onge
Anatolia EBA/MLBA
Dravidian
PC2 (12%)
PC1 (22.6%)
0.10
0.05
0.00
−0.05
−0.08
−0.06
−0.04
−0.02
0.00
0.02
0.04


Steppe
Namazga + S. Indian
Namazga + S. Indian + SteppeMLBA
Namazga + S. Indian + Xiongnu
TurkmenistanIA + Xiongnu
Anatolia
CHG
YamnayaKaragash_EBA
YamnayaAfanasievo_EBA
EHG
SidelkinoEHG_ML
MA1
Botai_CA
CentralSteppe_EMBA
Okunevo_EMBA
UstIda_BA
Kurma_EBA
Shamanka_EBA
UstIda_LN
Lokomotiv_EN
Shamanka_EN
Iran_N
Namazga_CA
Gujar
Iyer
Maratha
Pallan
Kadar
Irula
Gond
Paniya
Korwa
Steppe_MLBA
Yagnobi
Tarkalani
Yusufzai
Uthmankhel
GujratiBrahmin
Khatri
WestBengalBrahmin
Kohistani
Xiongnu_IA
Jamatia
Tripuri
ManipuriBrahmin
Tharu
Ho
Santal
Birhor
Onge
Turkmenistan_IA
Tajiks
Tajiks_Pamir
WHG
Anatolia_N
Anatolia_CA
Anatolia_EBA
Anatolia_MLBA
Anatolia_IA
Anatolia_Ottoman



kya
140.0
120.0
100.0
80.0
60.0
40.0
30.0
25.0
20.0
15.0
10.0
5.0
N
1.0e+02
4.0e+02
1.6e+03
6.4e+03
2.6e+04
CHG
ANE
EHG
AEA
MA1
KK1
Sidelkino EHG_ML
Yamnaya Karagash_EBA
40% (+0±8)
16% (+5±1)
49% (+3±3)
45% (+2±1)
Botai_CA
Okunevo EMBA
Central Steppe EMBA
Shamanka EN
Mbuti
Han


A
San A1b1a1a1
Dinka A1b1b2b1
San A1b1b2a
Dai C2
Uyghurs C2e1a1a
Buryats C2e1a1a
Nivhks C2b
Kalmyks C2b1c
Uyghurs C2b1c
Buryats C2b1c
Hazaras C2b1c
Hazaras C2b1c
Lokomotiv_EN C2b1a1
Kazakhs C2b1a1b
Koryak C2b1b2
Kazkahs C2b1b1
Karakalpaks.3 C2b1b1
TomskTatars C2b1b1
Kalmyks C2b1b1
Kalmyks C2b1b1
Dungan H1b2
Karakalpaks G2a2b2a1a1a2b
Turkmens G2a2b2a1a1a2a1
Anatolia_MLBA G2a2b1
French I1
Bichon I2a1a2a
Sardinian I2a1a1a1a1a
Aleutian I2a1b2a1a2
Loschbour I2a1b
VolgaTatars I2a1b2a1b1
Belarusian I2a1b2a1b1
Sardinian J1a2b
Namazga_CA J
KK1 J2a
Anatolia_EBA J2a
VolgaTatars J2a1
BR2 J2a1b
Namazga_CA J2a1
Anatolia_MLBA J2a1
Anatolia_MLBA J2a1
Kalash L1a2
Ust_Ishim NO
Shamanka_EN NO1
Han O2a2b1a1a
Han O1b1a1b
Dai O1b1a2b1a
Hazaras O1b1a2a1
Altais N
Botai_CA N
Shamanka_EN N
Shamanka_EN N1
Lokomotiv_EN N1c
Komi N1c2
Khanty N1c2
Shamanka_EN N1c2
Shamanka_EN N1c2
Shamanka_EN N1c2
Bashkirs N1c1a1
UstIda_LN N1c1
Even N1c1a1
Yakut N1c1a1
Evenk N1c1a1
Khanty N1c1a1a2b
Komi N1c1a1a
Uzbeks N1c1a1a
Buryat N1c1a1a
Buryat N1c1a1a
Buryat N1c1a1a
MA1 R
Karakalpaks R2a
TomskTatars R1b1
Botai_CA R1b1a1
Teleuts R1b1a1
French R1b1a2a1a1d2
YamnayaKaragash_EBA R1b1a2a2c1
Avar R1b1a2a2
Okunevo_EMBA R1b1a2a2
IB_Tatars R1a1a1b1a2a
Kazakhs R1a1a1b1a2a
Khanty R1a1a1b1a2b3
Russian R1a1a1b1a2b3
Turkmenistan_IA R1a1a1b2
Altaian R1a1a1b2
Kyrgyz R1a1a1b2
Kyrgyz R1a1a1b2
Altaian R1a1a1b2
Teleuts R1a1a1b2
Bashkirs R1a1a1b2a2a
Tajiks R1a1a1b2a2a
Altai R1a1a1b2
Bashkirs R1a1a1b2
Uyghurs R1a1a1b2a1b
Tajiks R1a1a1b2a1a
Kurma_EBA Q1a
Okunevo_EMBA Q1a
Greenland Q1a1a
Athabascan Q1a1
Okunevo_EMBA Q1a1b1
Uzbeks Q1a1b1a
UstIda_LN Q1a2
UstIda_EBA Q1a2
Okunevo_EMBA Q1a2
Okunevo_EMBA Q1a2
Tajiks Q1a2
Okunevo_EMBA Q1a2b
Dungan Q1a2b
Turkmens Q1a2a
Shamanka_EBA Q1a2a
Shamanka_EBA Q1a2a
Shamanka_EBA Q1a2a
Shamanka_EBA Q1a2a
Shamanka_EBA Q1a2a
UstIda_EBA Q1a2a
UstIda_EBA Q1a2a
UstIda_EBA Q1a2a1
Okunevo_EMBA Q1a2a1
North America Q1a2a1b
Mayan Q1a2a1b
Pima Q1a2a1b
Yukpa Q1a2a1a1
Kennewick Q1a2a1a
Athabascan Q1a2a1a1
Aymara Q1a2a1a1
Karitiana Q1a2a1a1
Karitiana Q1a2a1a1
Yeniseian Q1a2a1c
Shamanka_EBA Q1a2a1c
Okunevo_EMBA Q1a2a1c
Okunevo_EMBA Q1a2a1c
Okunevo_EMBA Q1a2a1c
IB_Tatars Q1a2a1c
Yeniseian Q1a2a1c
C
G
J
NO
R
Q
Shamanka_EN
Lokomotiv_EN
Botai_CA
Namazga_CA
YamnayaKaragash_EBA
UstIda_LN
Okunevo_EMBA
Kurma_EBA
Anatolia_EBA
Anatolia_MLBA
UstIda_EBA
Shamanka_EBA
Turkmenistan_IA
B
San L0d1c1a1a
San L0d1b2b2c1
Mbuti L0a2b
Mbuti L0a2b
Mandenka L1b1a
Yoruba L1b1a
Mandenka L2c3a
Dinka L2c1
Yoruba L2a1a2b
Teleuts Z1a
Botai_CA Z1a
Kalmyks Z3
Mixe C1b1
Pima C
IB_Tatars C5b1a
Yakut C4a2a1a
Kazakhs C4a2a1
Buryats C4a1a
Buryat C4a1a3b
Altai C4b3
CentralSteppe_EMBA C4
CentralSteppe_EMBA C4a1a4a
Shamanka_EN C4
Shamanka_EN C4
Okunevo_EMBA C5c
Okunevo_EMBA C5c
Okunevo_EMBA C5c
Okunevo_EMBA C5c
UstIda_EBA C4a1a3
UstIda_EBA C4a1a3
Shamanka_EBA C4a1a3
Evenk C4a2a1b
Shamanka_EBA C4a2a1
Shamanka_EBA C4a2a1
Shamanka_EBA C4a2a1
Shamanka_EN C4
Buryats G2a2
Dungan G2b1a
Shamanka_EN G2a1
Shamanka_EN G2a1
Shamanka_EBA G2a1
Shamanka_EN G2a1
Buryat D4p
Teleuts D4b2b
Karitiana D1e
Karitiana D1
Dai D4
Lokomotiv_EN D4j
Shamanka_EN D4j
UstIda_EBA D4j4
Shamanka_EN D4e1
Shamanka_EN D4e1
Aleutian D2a
North Clovis D4h3a7
Lokomotiv_EN D4
Anatolian_Ottoman D4j
Khanty D4l2
Okunevo_EMBA D4
UstIda_LN D4j
Buryat D4p
Lokomotiv_EN D4
Even D3
Uzbeks D5b1b1
Kalmyks F1a1
Dungan F1a1c3
Turkmens F1b1
Kyrgyz F1b1b
Kurma_EBA F1b
Kurma_EBA F1b
Shamanka_EBA F1b1b
Shamanka_EN F1b1
Kalmyks F1b
Bashkirs F2f
Anatolia_MLBA U1a
MA1 U
Yeniseian U5a1d2
SidelkinoEHG_ML U5a2
Saami U5b1b1a3
French U5b1c1a
Karakalpaks U4a1a
Yeniseian U4a1
VolgaTatars U2e2a1d
Kalash U2e1h
Turkmens U2c1
TomskTatars U8a1a1
UstIda_LN R1b1
Botai_CA R1b1
YamnayaKaragash_EBA R1a1a
Anatolia_MLBA J1c
Anatolia_EBA J1c10a
Namazga_CA J1
Anatolia_MLBA J2b1
Tajiks J1d2
Sardinian J2b1a
Altais T1a1
French T1a
Uyghurs T2b
Khanty T2d1b1
Komi T2d1b1
Turkmenistan_IA T2c1a
UstIda_EBA H2a2a
Russian H2a2a
Anatolia_MLBA H6a1b2e
Belarusian H5m
Komi H11a1
Okunevo_EMBA H15b1
Okunevo_EMBA H6a1b
Okunevo_EMBA H6a1b
Okunevo_EMBA H6a1b
Okunevo_EMBA H6a
Tajiks H6b
Hazaras HV
Kyrgyz HV1a2
IB_Tatars H2b
TomskTatars H49a
Sardinian H3u
Uzbeks H78
VolgaTatars H4a1
Khanty H1e5
Russian H10g
Anatolia_EBA H
Kazakhs K1b2a2
Botai_CA K1b2
Anatolia_MLBA K1a
Anatolian_Ottoman K
Namazga_CA W3a2
Anatolia_EBA W5
Okunevo_EMBA A
Okunevo_EMBA A
UstIda_LN A
Lokomotiv_EN A
Athabascan A2
Mayan A2w1
Yukpa A2af1a1
Athabascan A2
UstIda_LN A2
Karakalpaks A
Bashkirs A1a
Uyghurs A11b
Karakalpaks.3 A11a
Han A5b1b
Koryak A8
Okunevo_EMBA A8a1
Okunevo_EMBA A8a1
Okunevo_EMBA A8a1
Okunevo_EMBA A8a
Okunevo_EMBA A8a
Okunevo_EMBA A8a
Okunevo_EMBA A8a
Z
C
G
D
F
U
R
J
T
H
K
W
A
SidelkinoEHG_ML
Shamanka_EN
Lokomotiv_EN
Botai_CA
Namazga_CA
YamnayaKaragash_EBA
UstIda_LN
Okunevo_EMBA
Kurma_EBA
Anatolia_EBA
Anatolia_MLBA
CentralSteppe_EMBA
UstIda_EBA
Shamanka_EBA
Turkmenistan_IA
Anatolian_Ottoman

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 Zacho1, 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, Andrey 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*

*correspondence to: rd@sanger.ac.uk and ewillerslev@snm.ku.dk

This PDF file includes:

Table of Contents
Supplementary Text
Figs. S1 to S37
Tables S1 to S2, Tables S4 to S13, and Tables S15 to S17
Captions for Table S3 and S14

Other Supplementary Materials for this manuscript includes the following:

Tables S3 and S14 as separate spreadsheets.

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

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

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 (not shown), 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 $.49 + .51 \cdot .4 \approx 0.69$ 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 South Asians (*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.

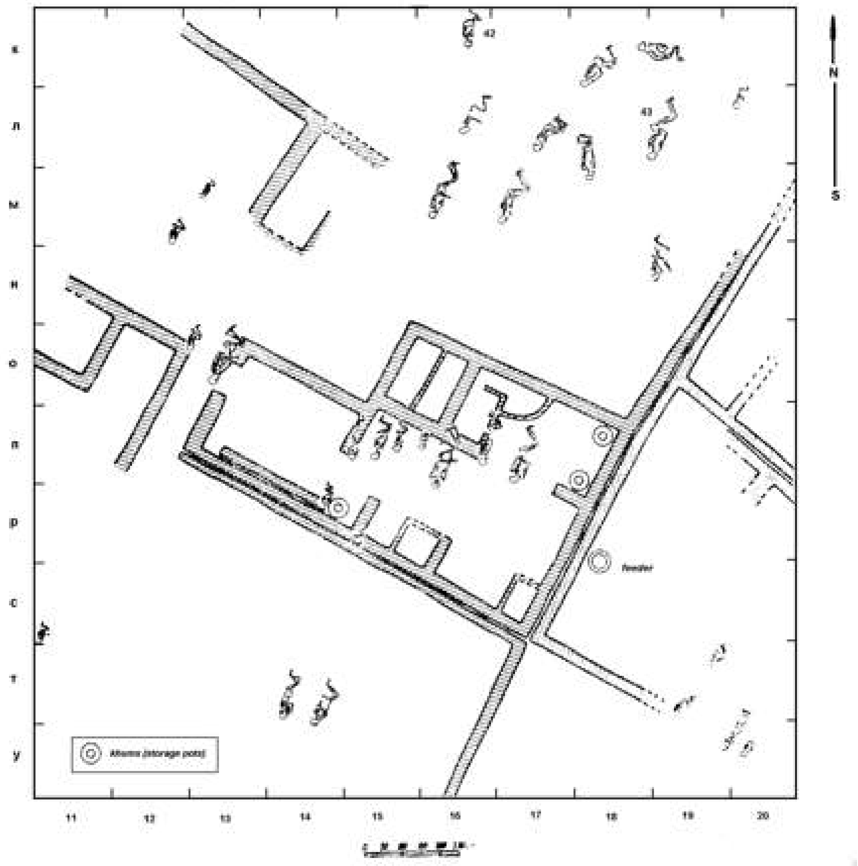


Fig. S1.
A plan of the excavation illustrates the burials of the skeletons (Namazga).

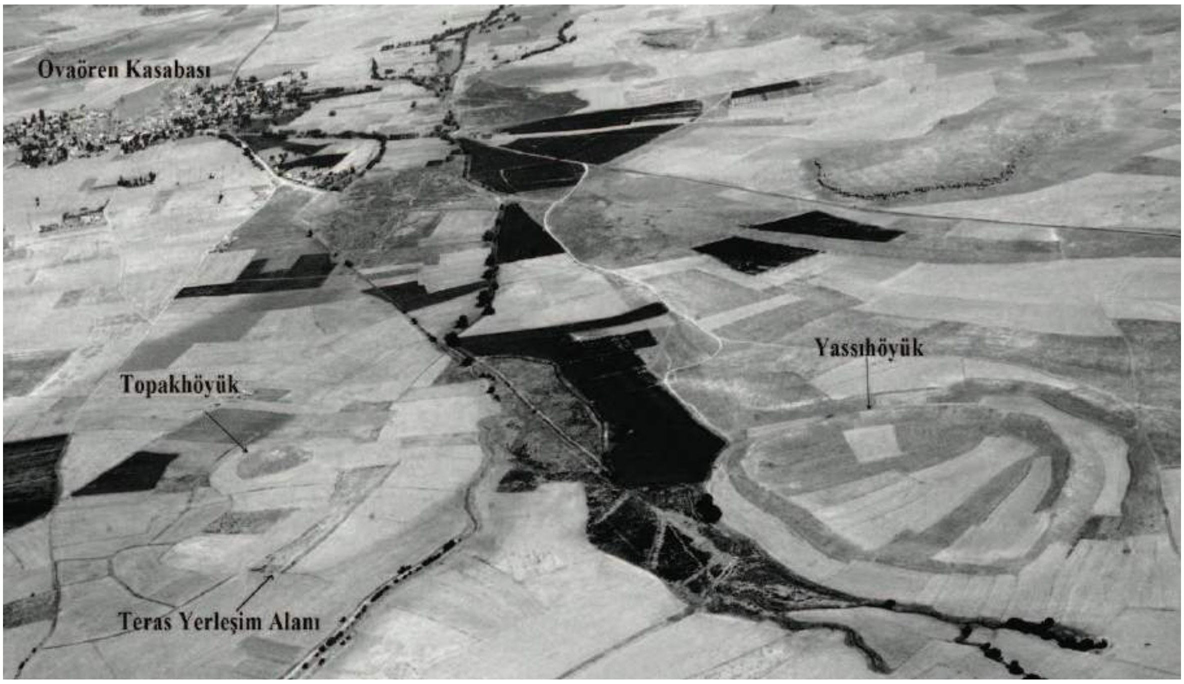


Fig. S2.
The three areas of Ovaören – Yassıhöyük, Topakhöyük, and the Terrace (Teras).

Fig. S3.

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

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

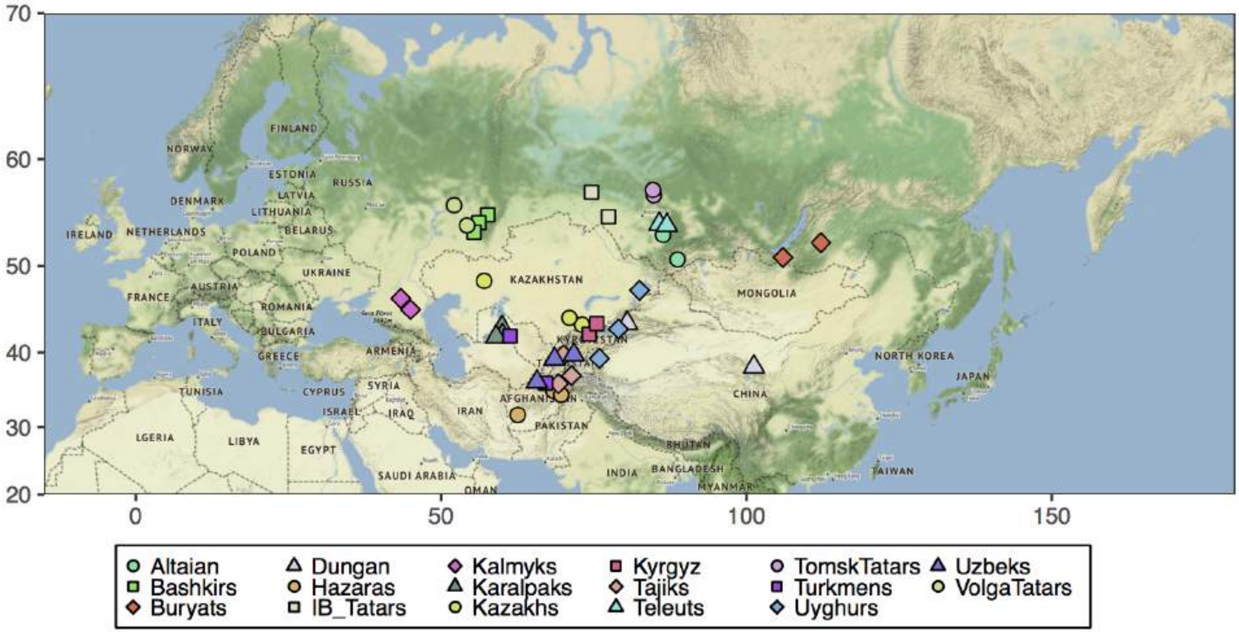


Fig. S5.

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

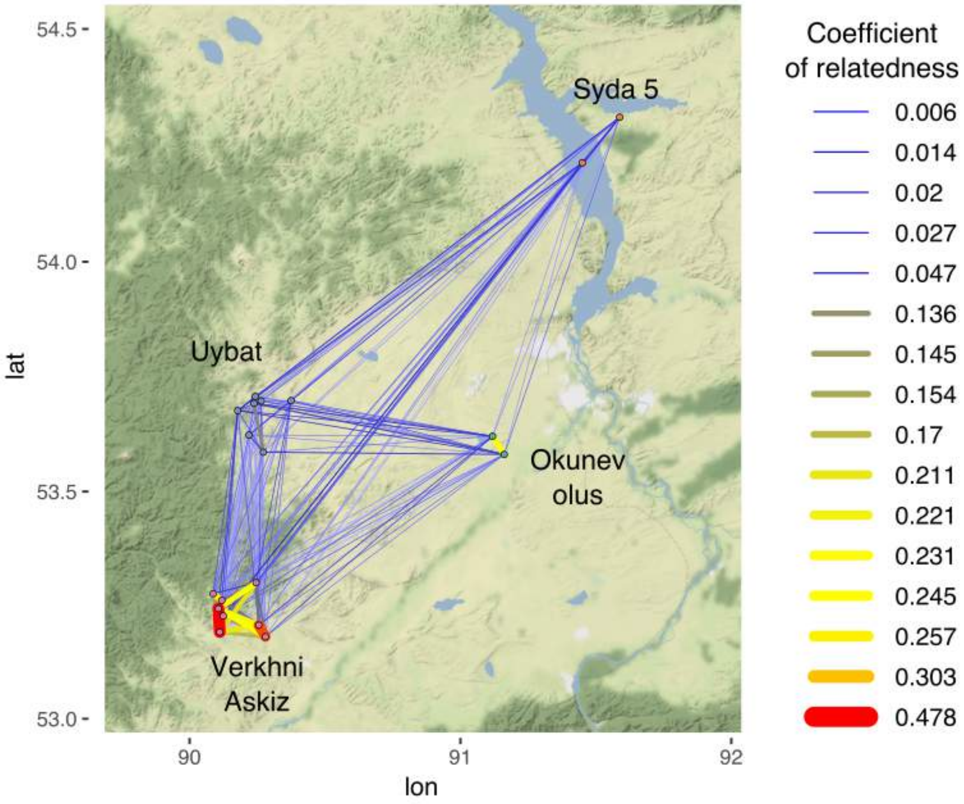


Fig. S6.

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.

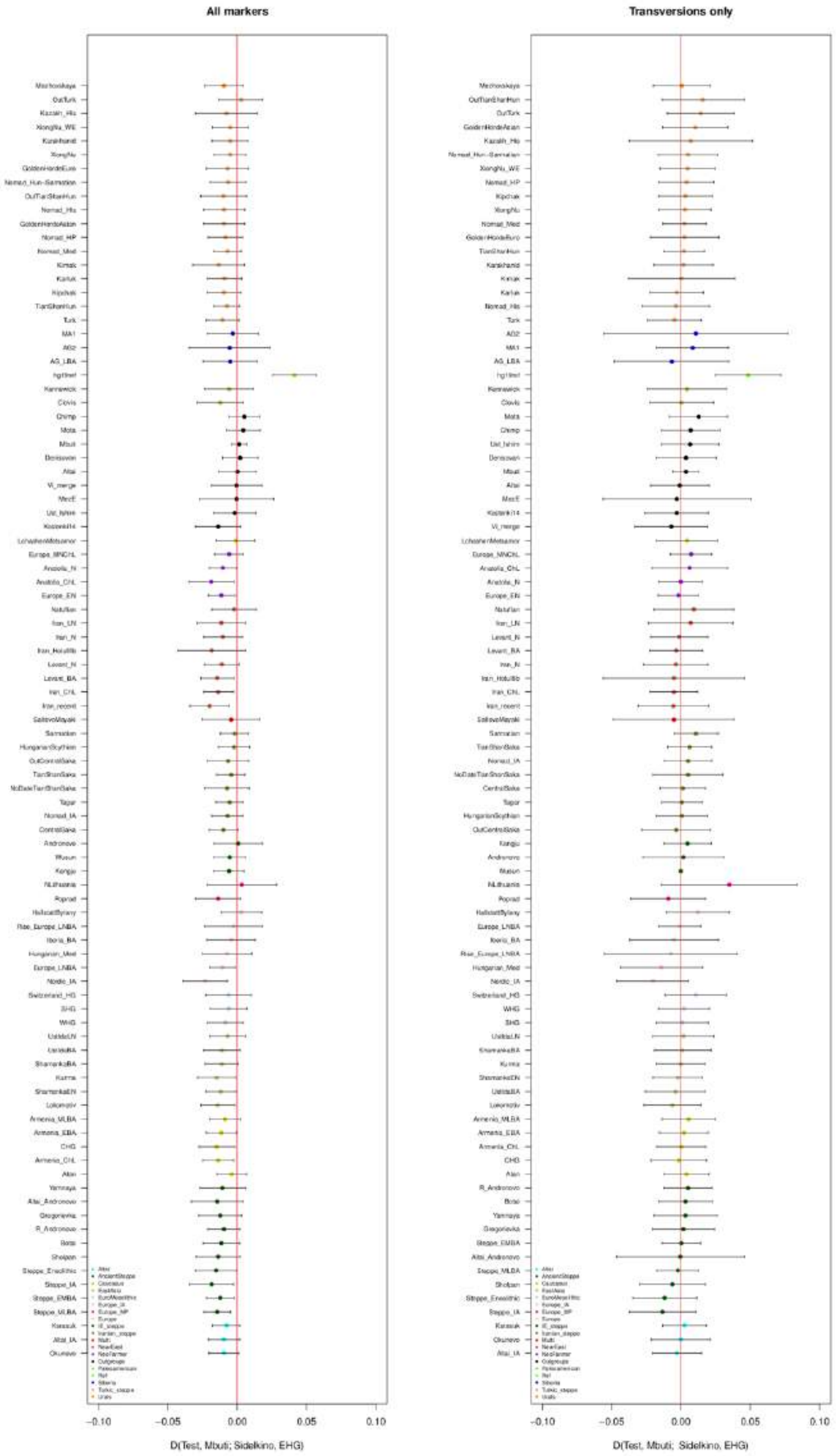


Fig. S7.

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

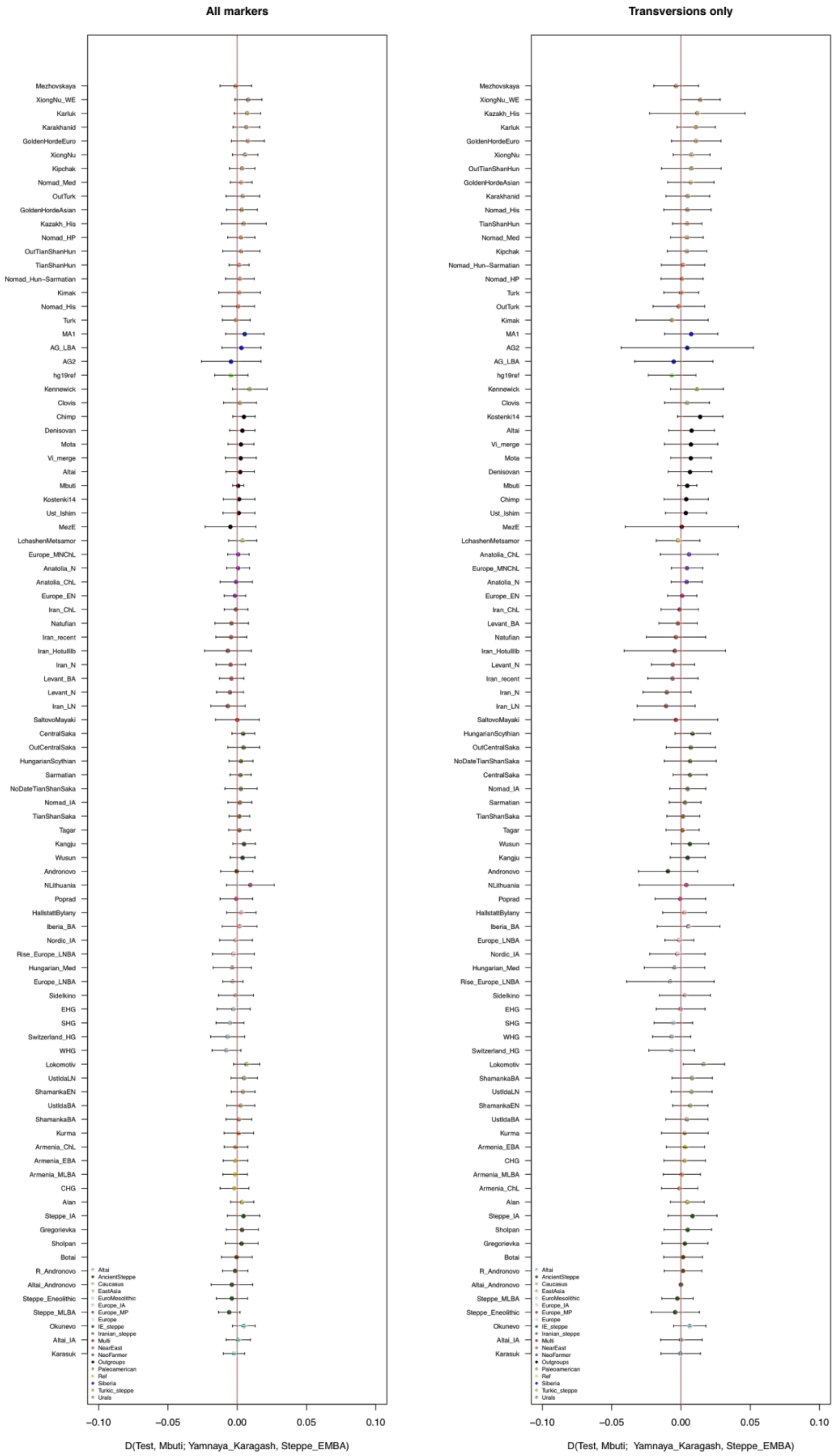


Fig. S8.

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

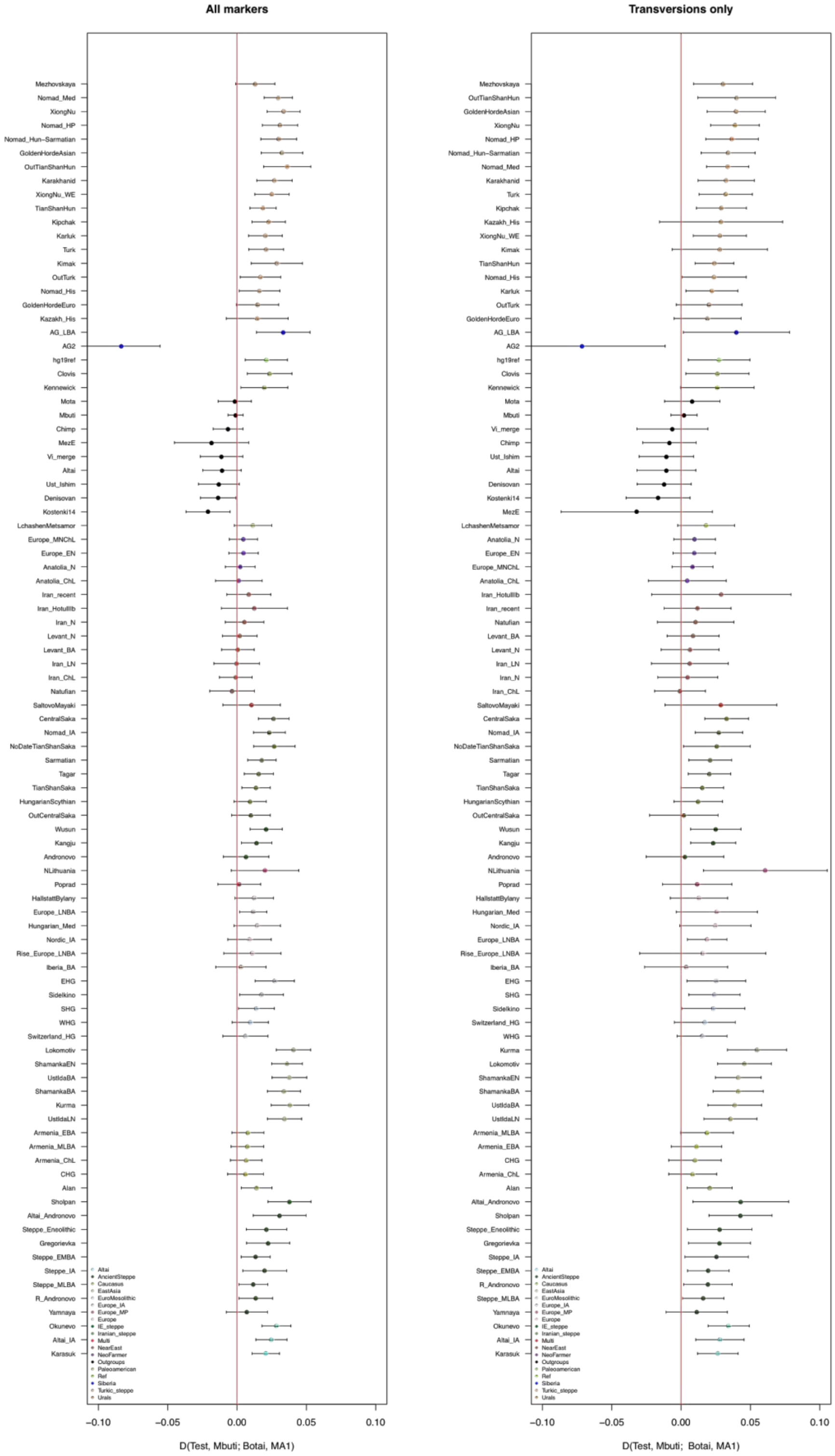


Fig. S9.

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

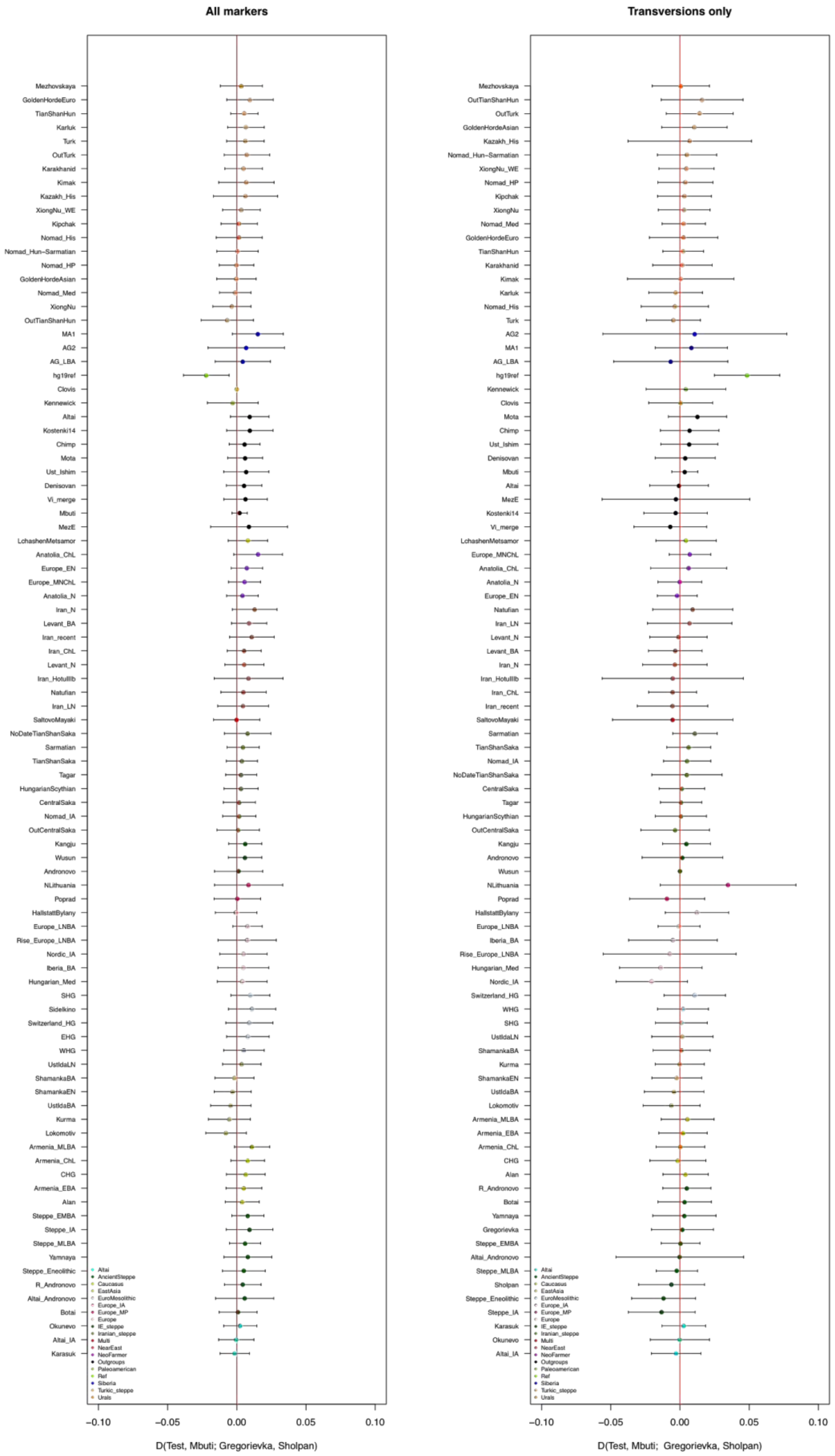


Fig. S10.

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

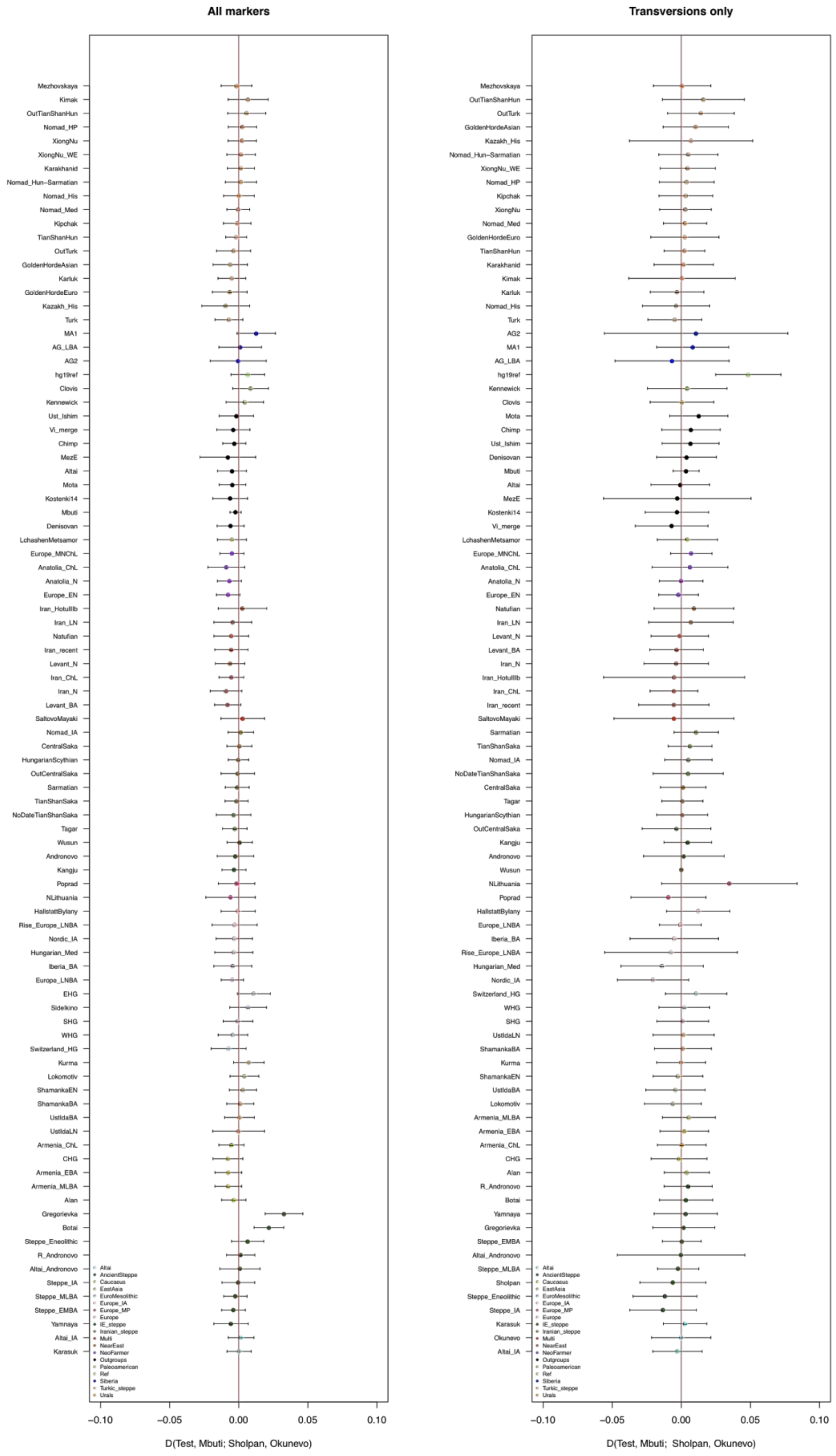


Fig. S11.

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

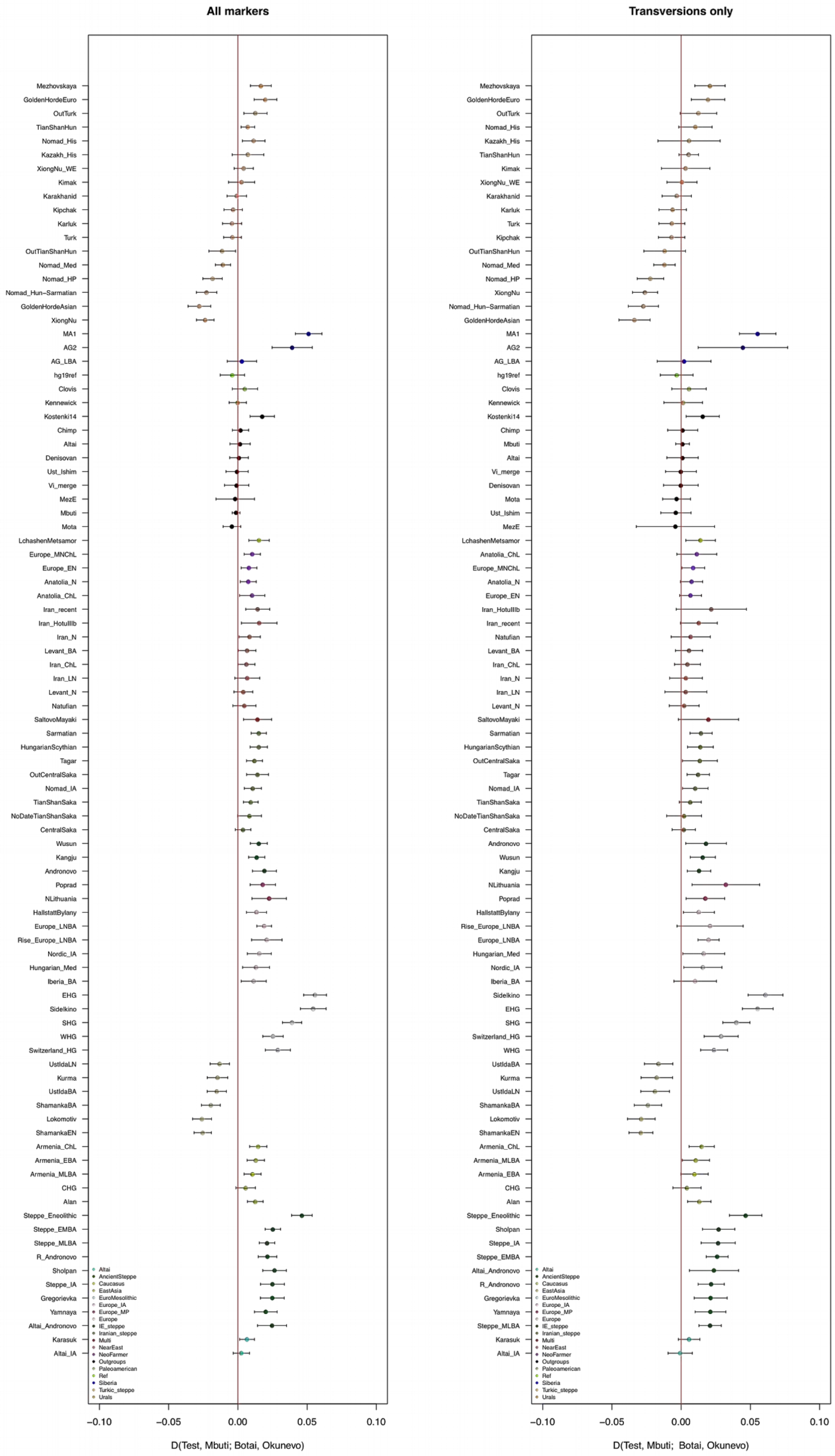


Fig. S12. D-statistics test of the form D(Test, Mbuti; Botai, Okunevo).

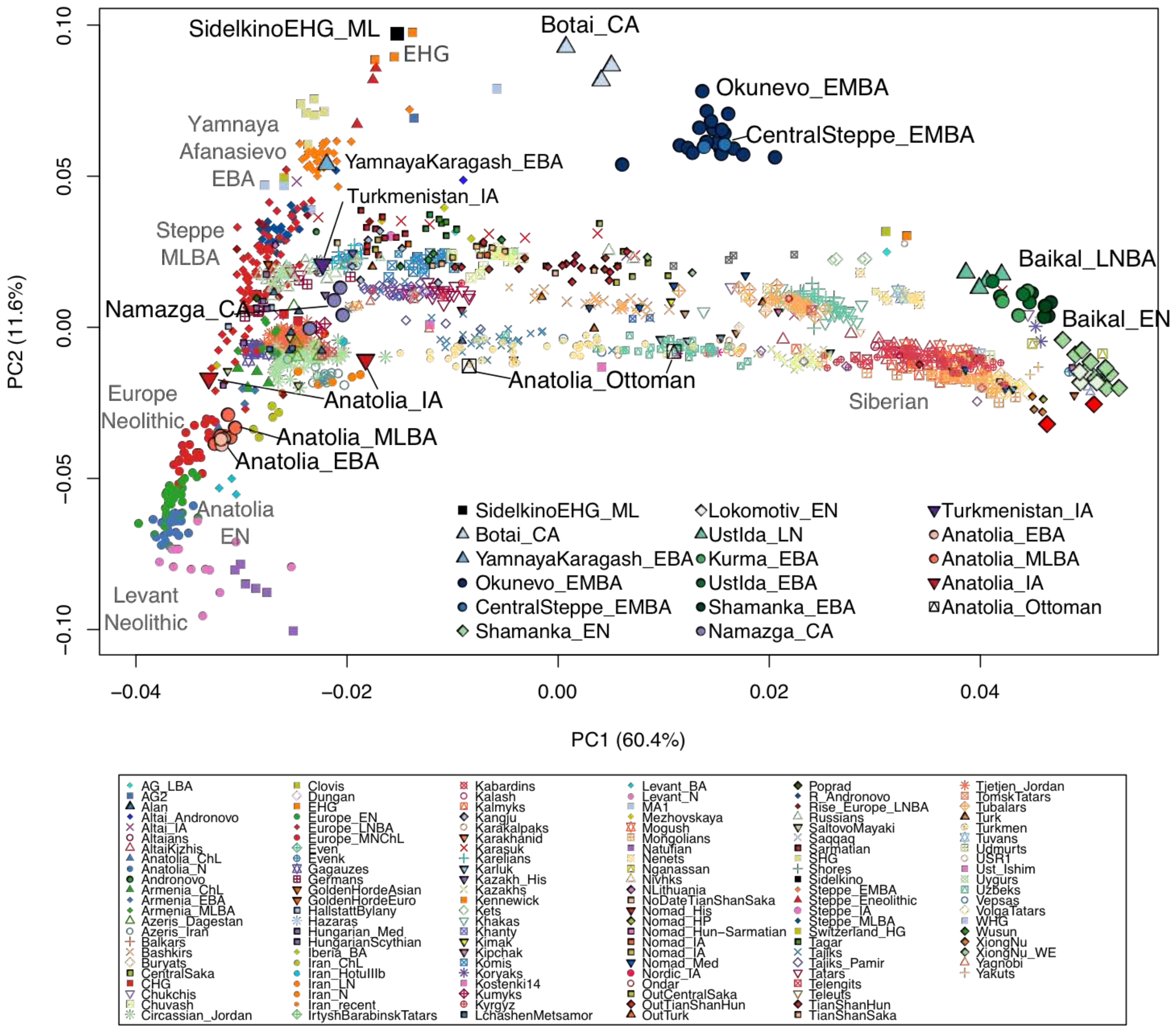


Fig. S13.

Principal Component Analysis estimated with ancient and modern Eurasians.

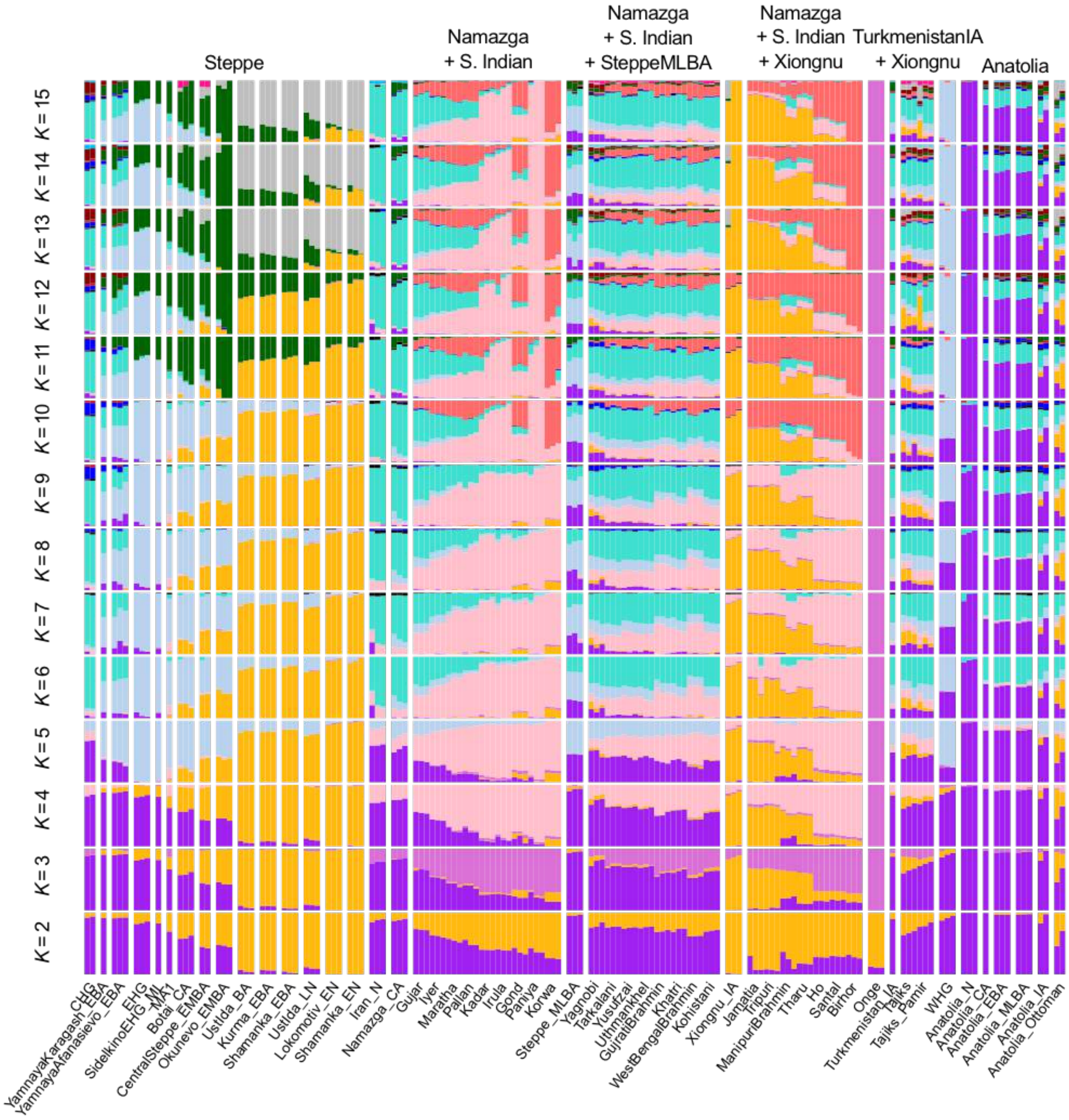


Fig. S14.

ADMIXTURE analysis for K = 2–15.

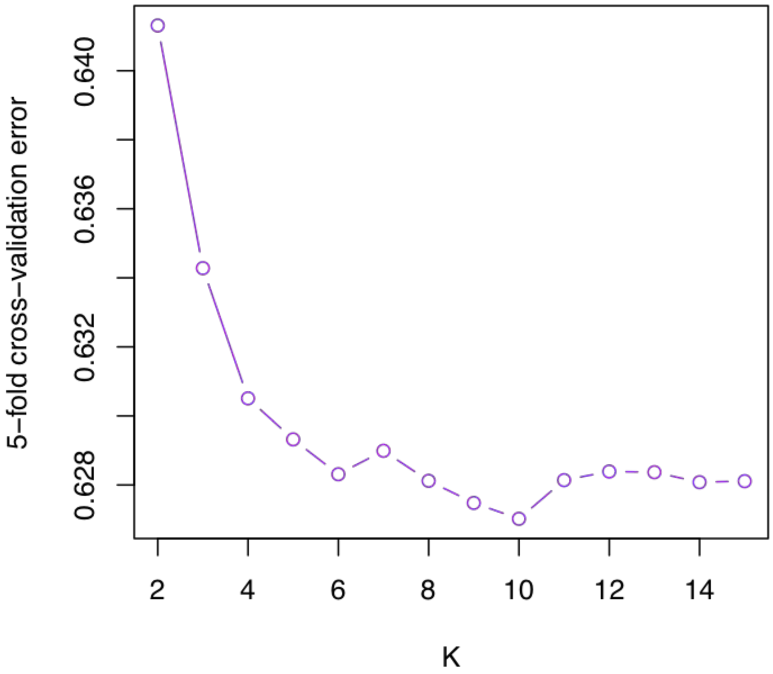


Fig. S15.
Cross-validation errors for the ADMIXTURE analysis.

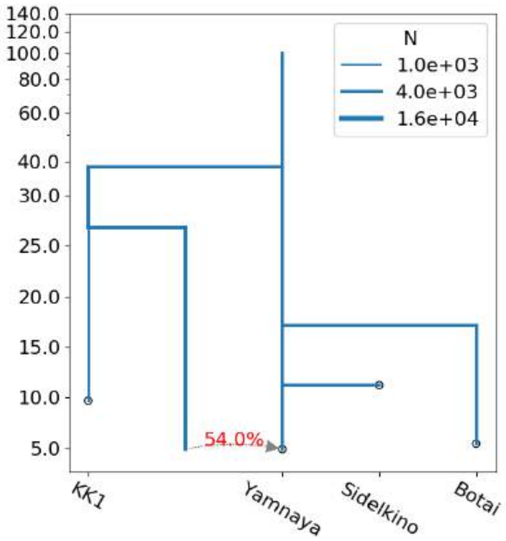


Fig. S16.

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.

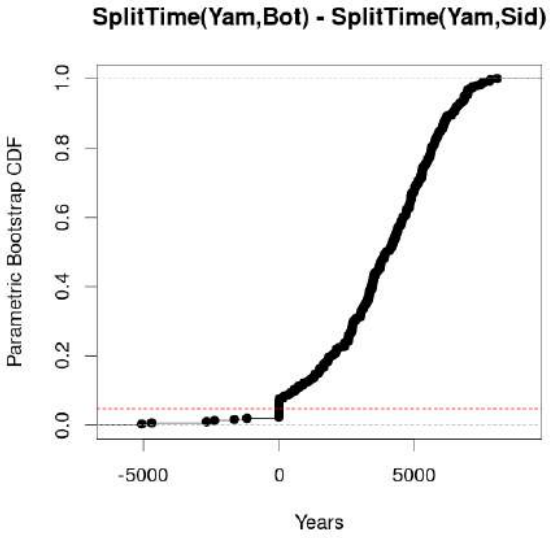


Fig. S17.

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).

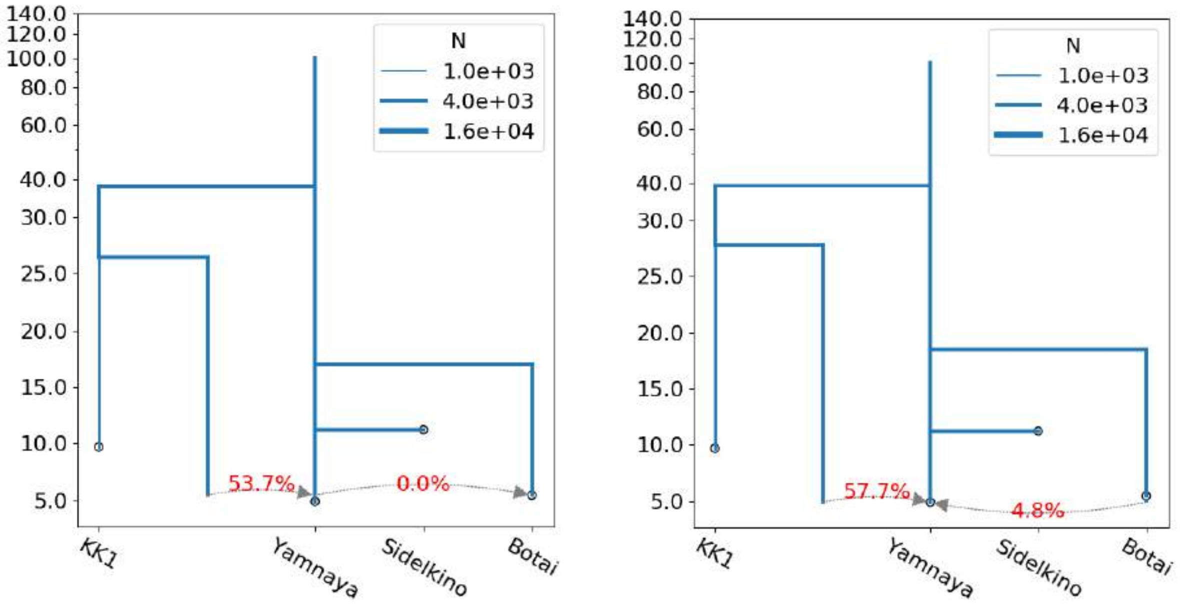


Fig. S18.

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).

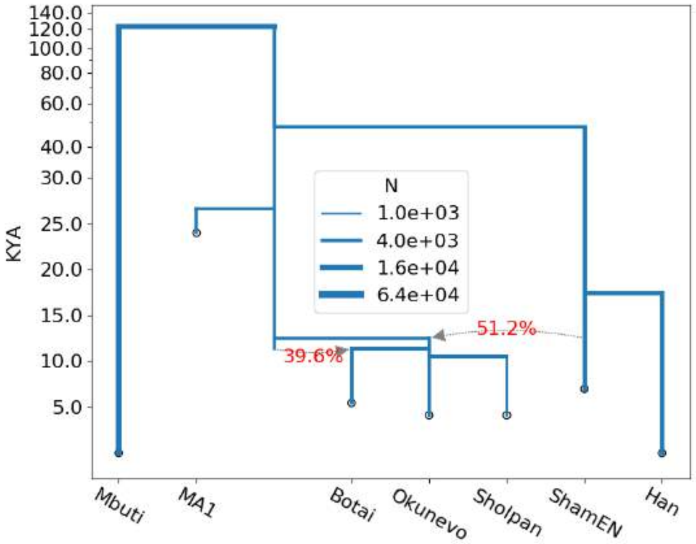


Fig. S19.

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.

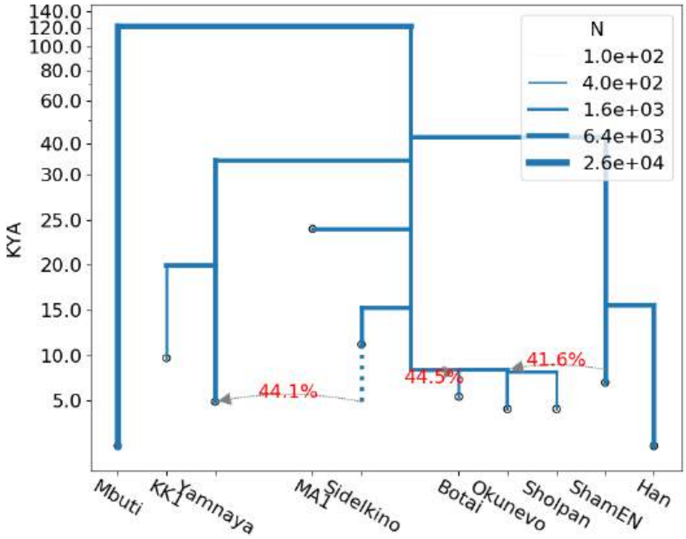


Fig. S20.

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

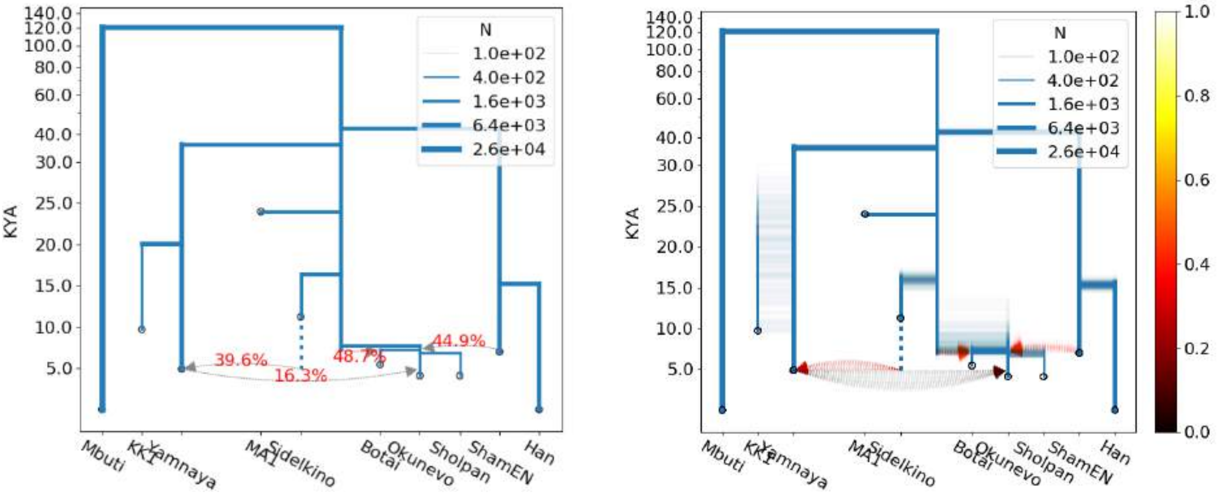


Fig. S21.

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.

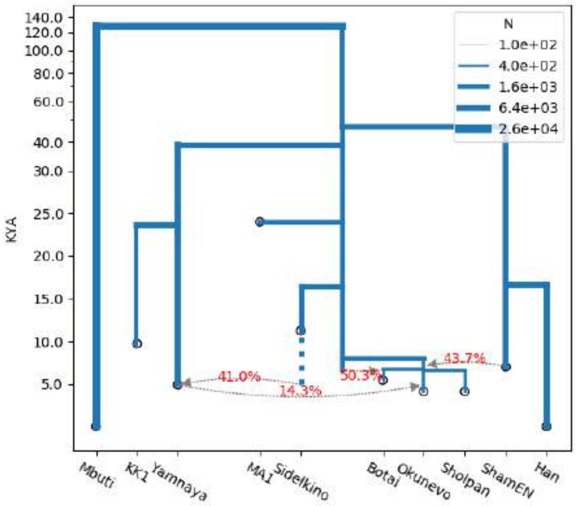


Fig. S22.

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.

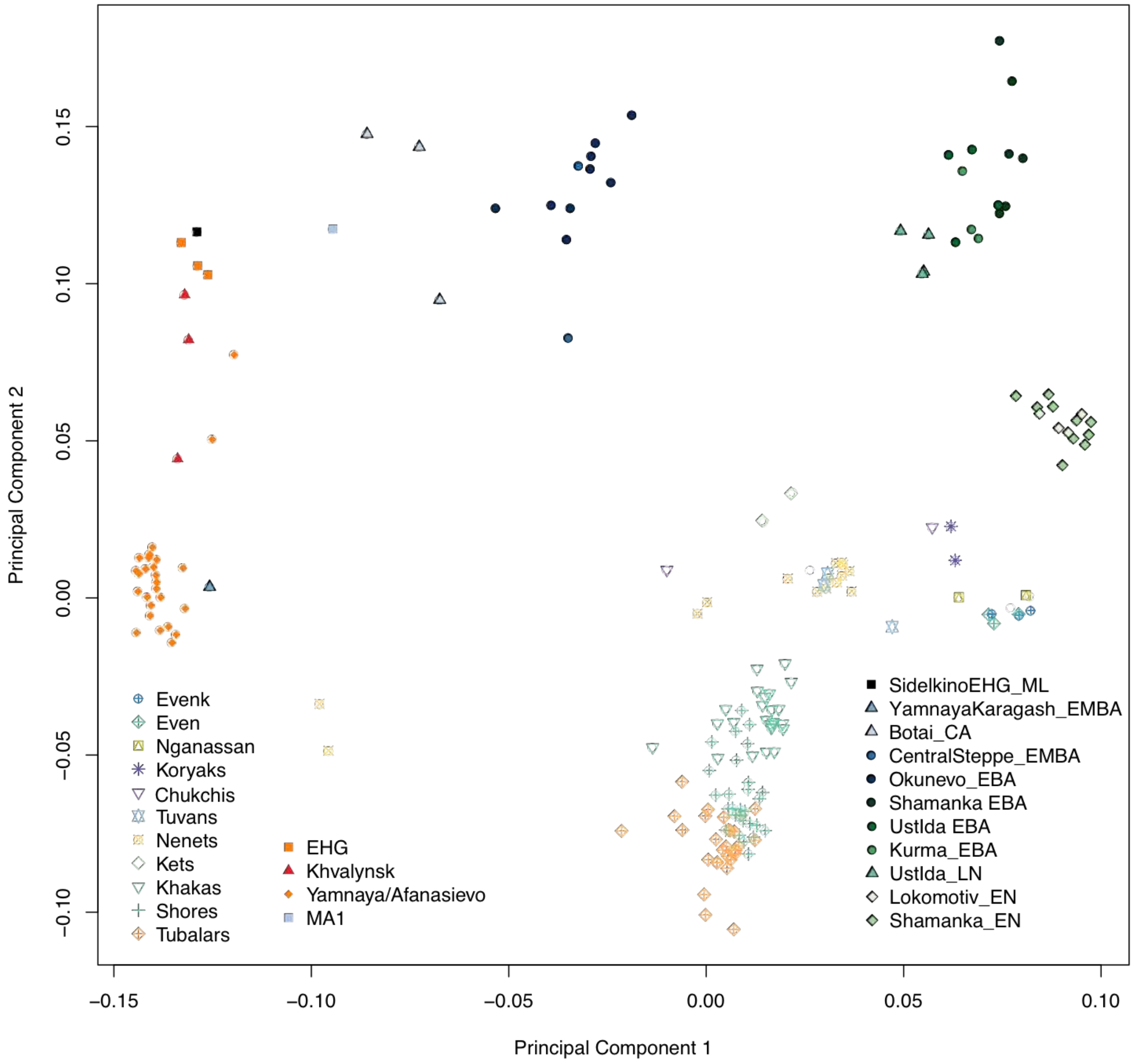


Fig. S23.

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

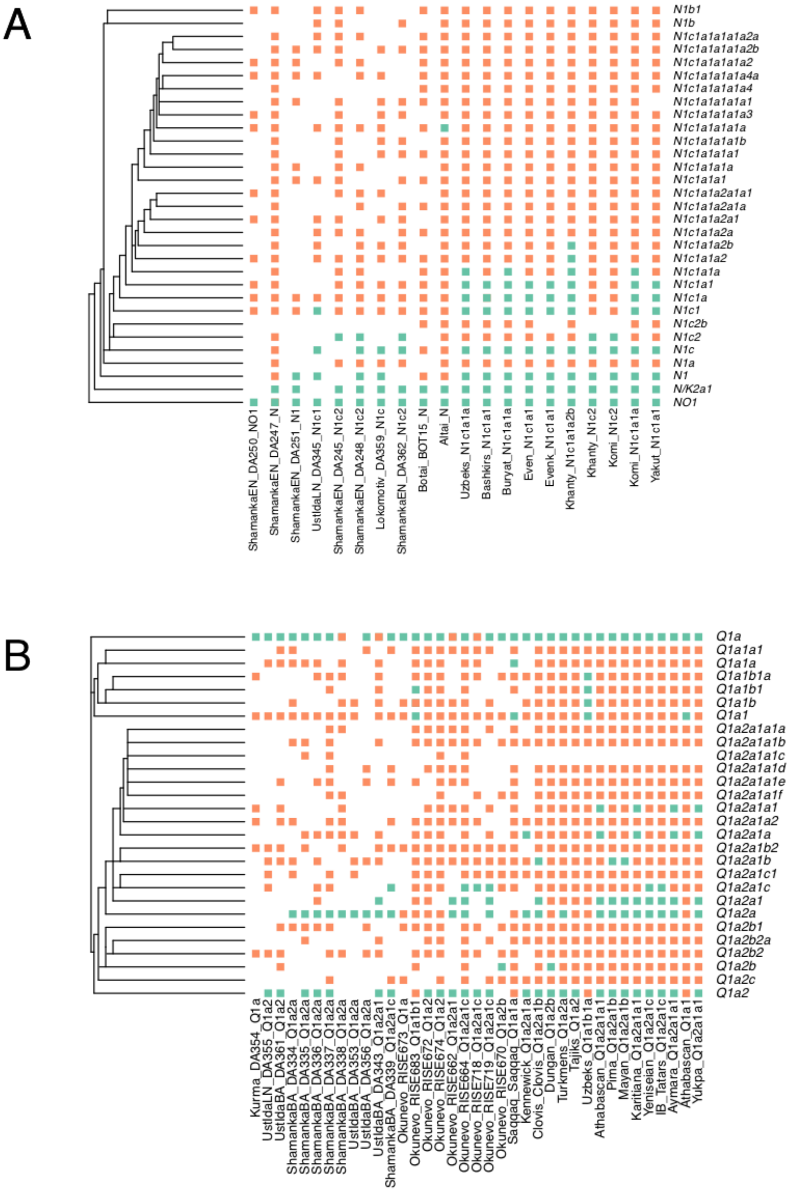


Fig. S24.

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.

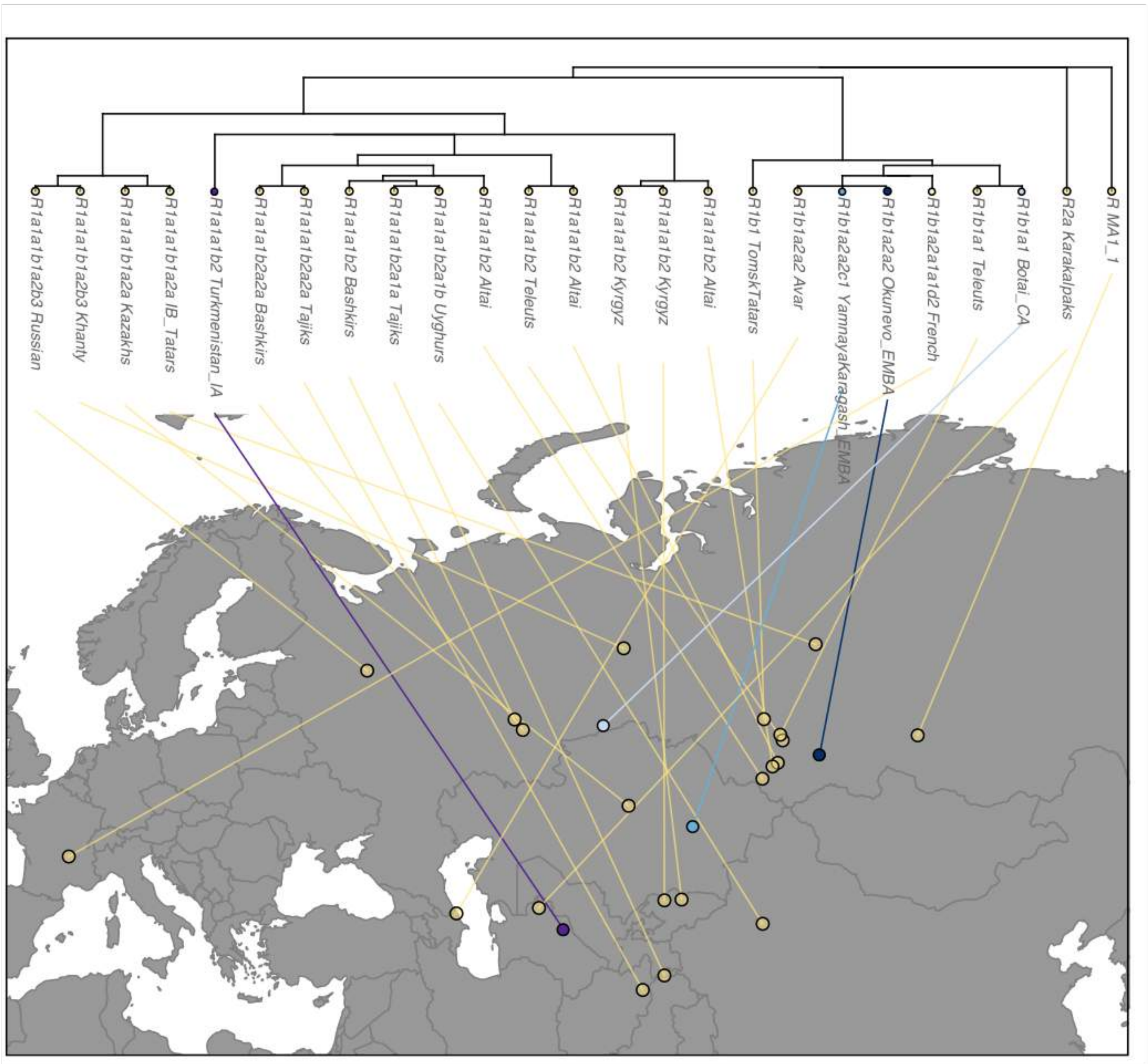


Fig. S25.

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

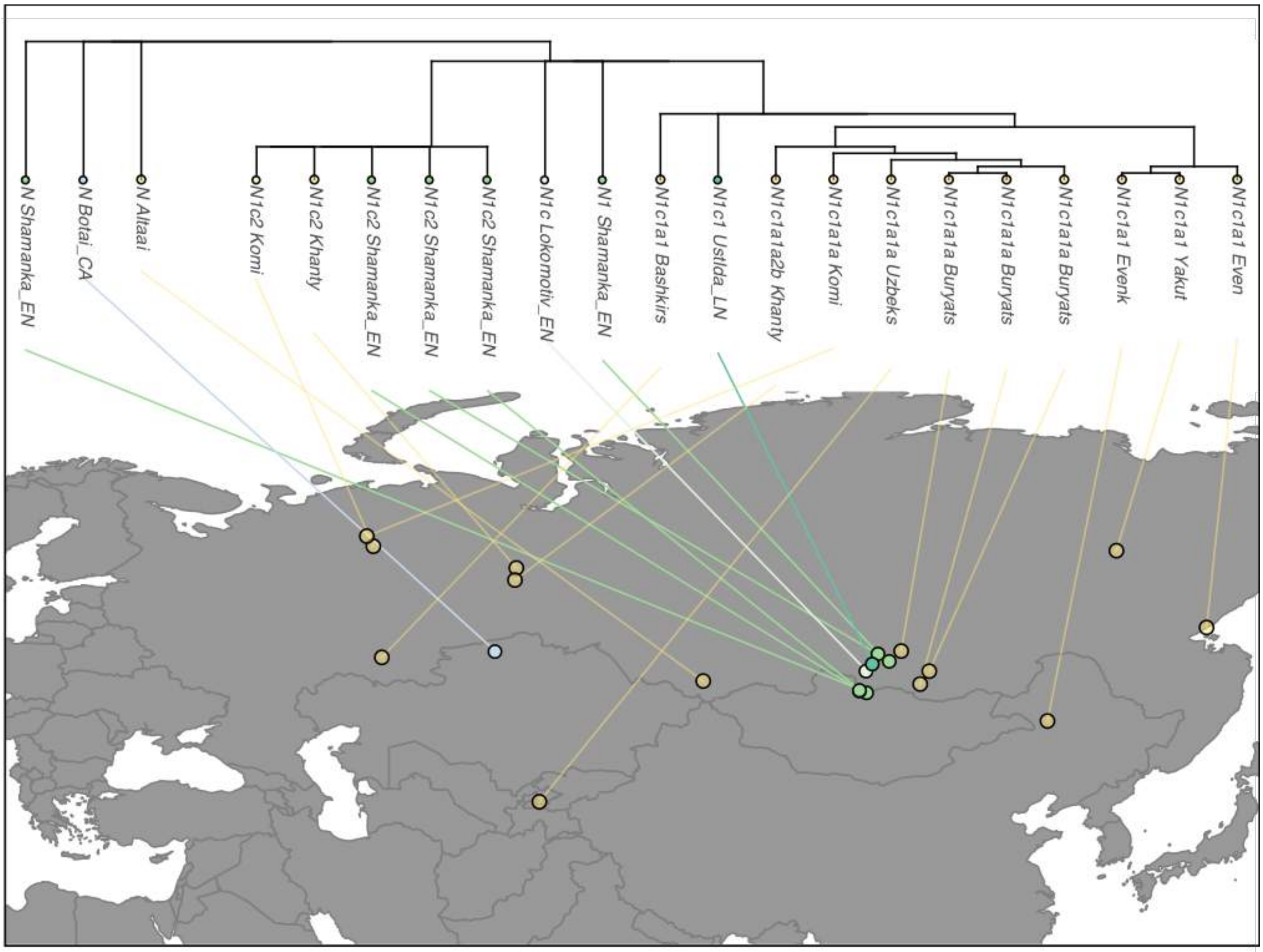


Fig. S26.

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

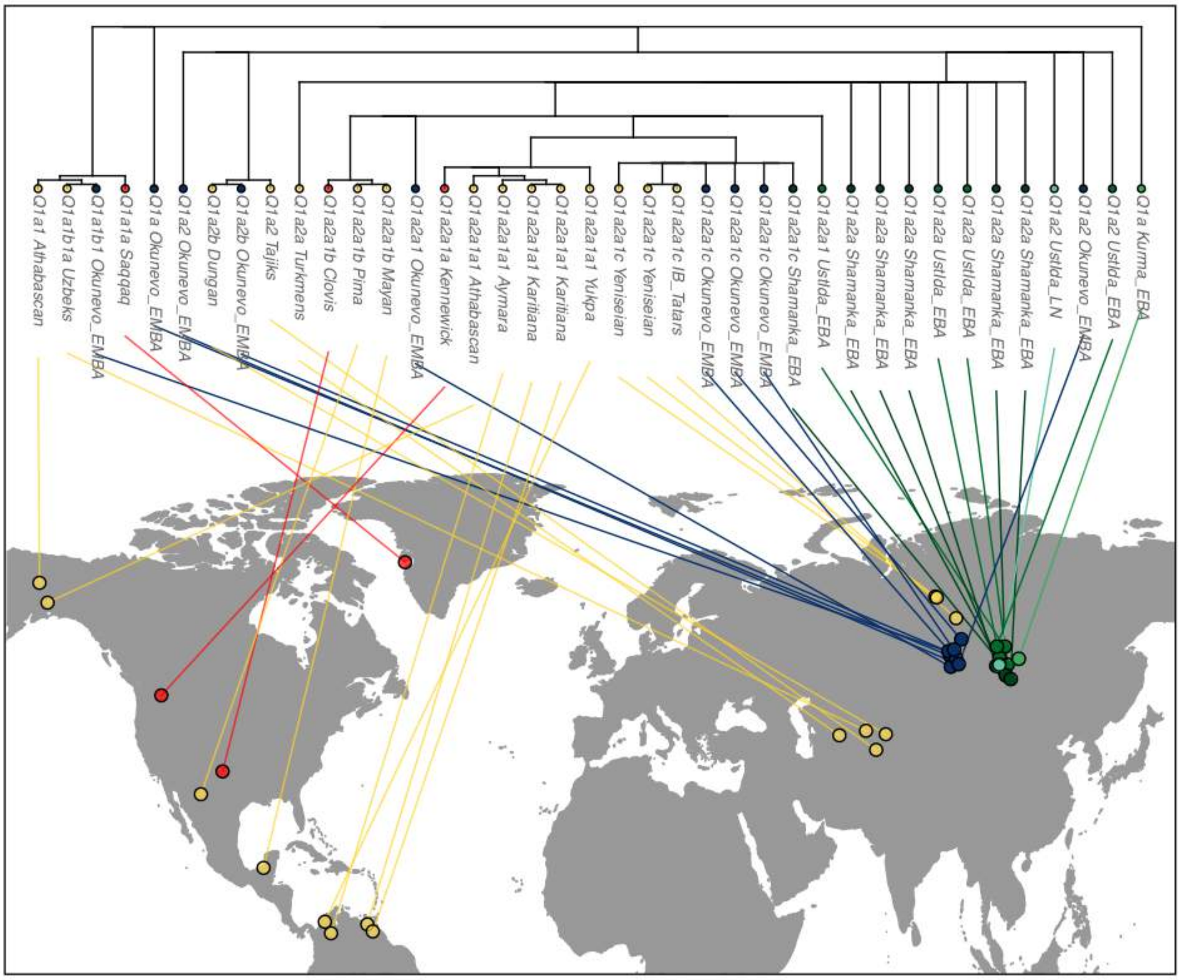
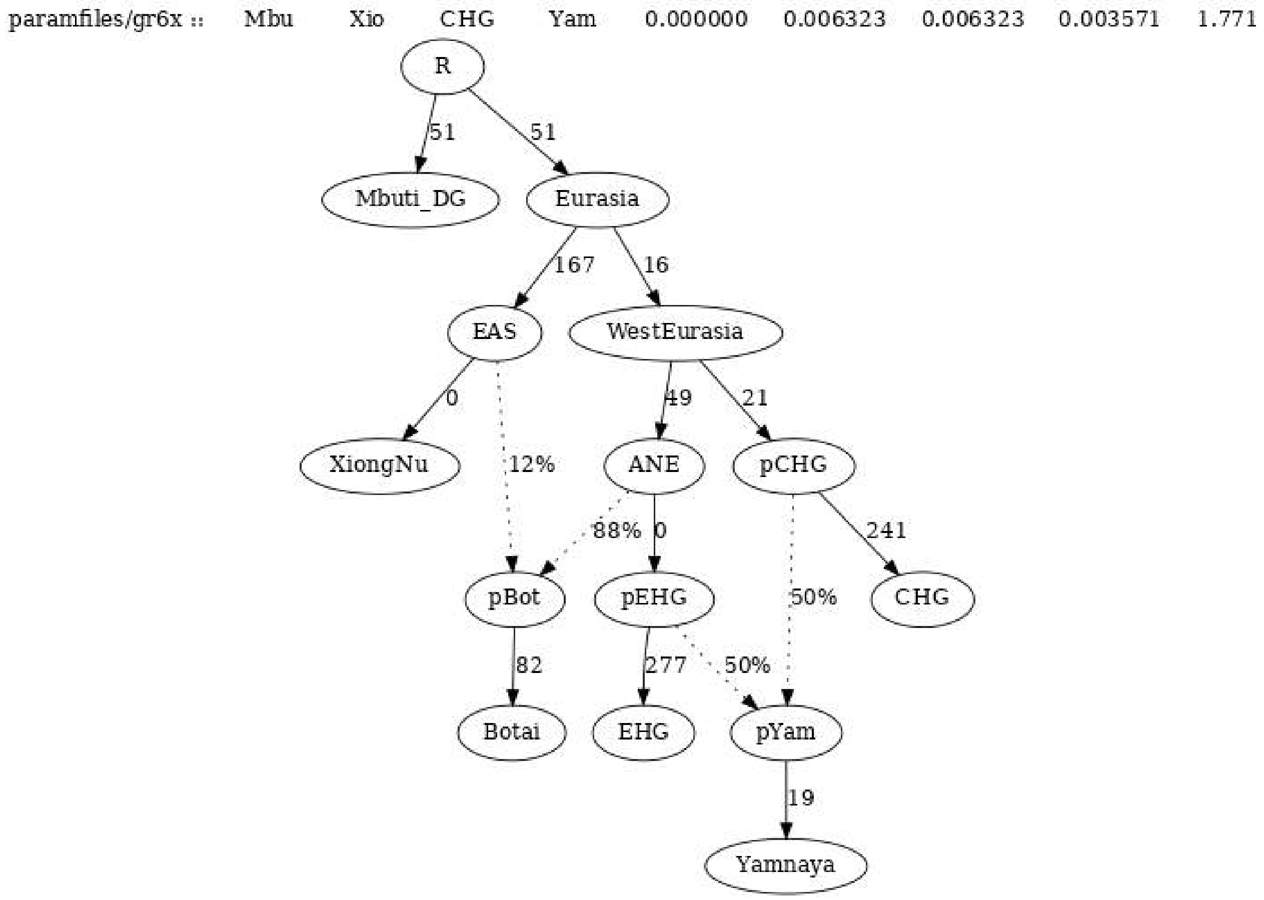


Fig. S27.
Geographical location of ancient samples belonging to major clade Q of the Y-chromosome.



```
paramfiles/gr6x ::    Mbu    Xio    CHG    Yam    0.000000    0.006323    0.006323    0.003571    1.771
```

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).

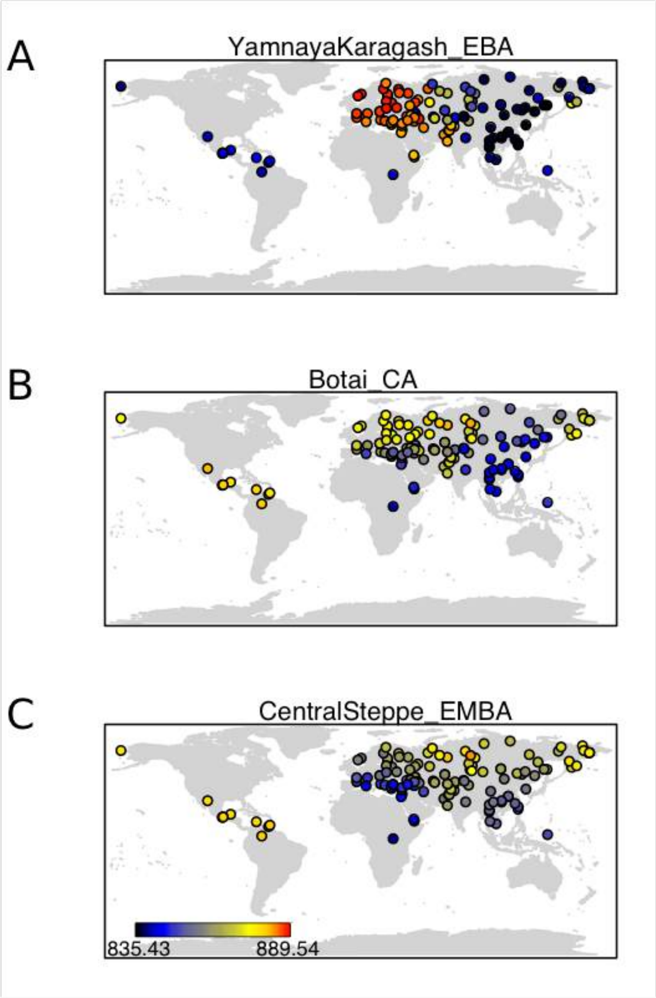


Fig. S29.

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

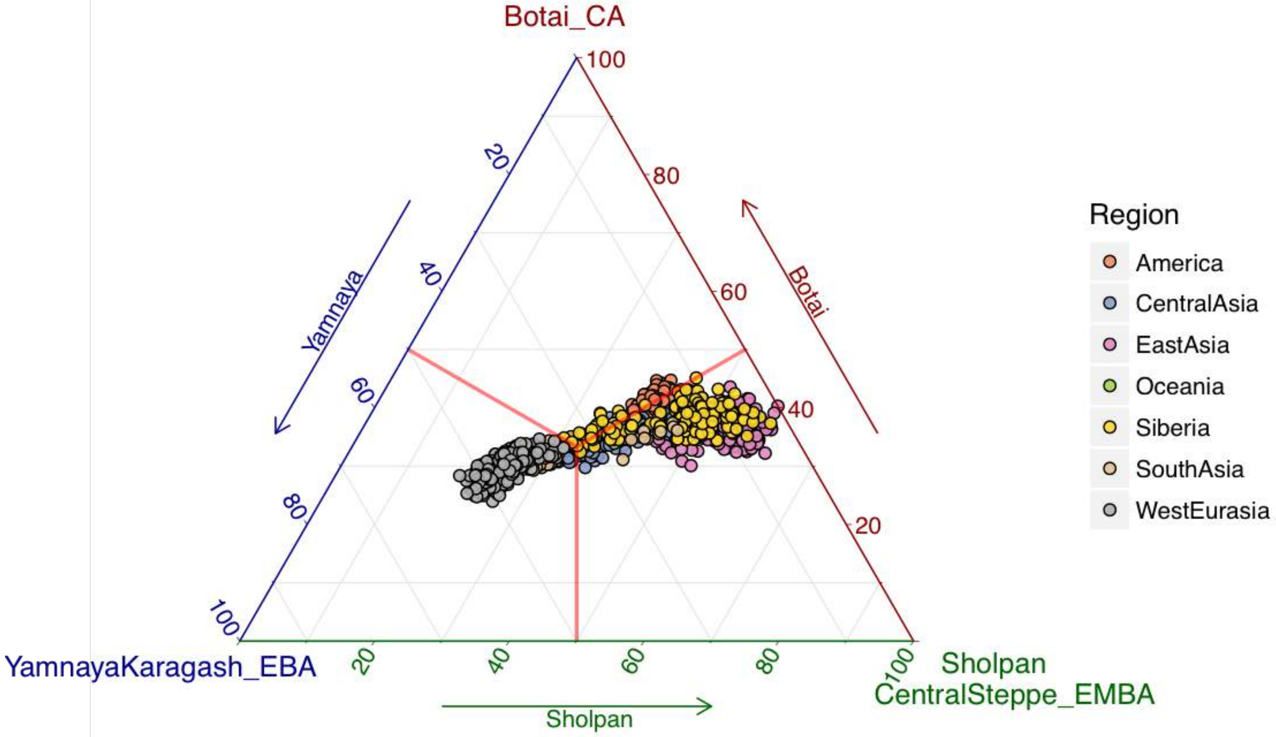


Fig. S30.

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

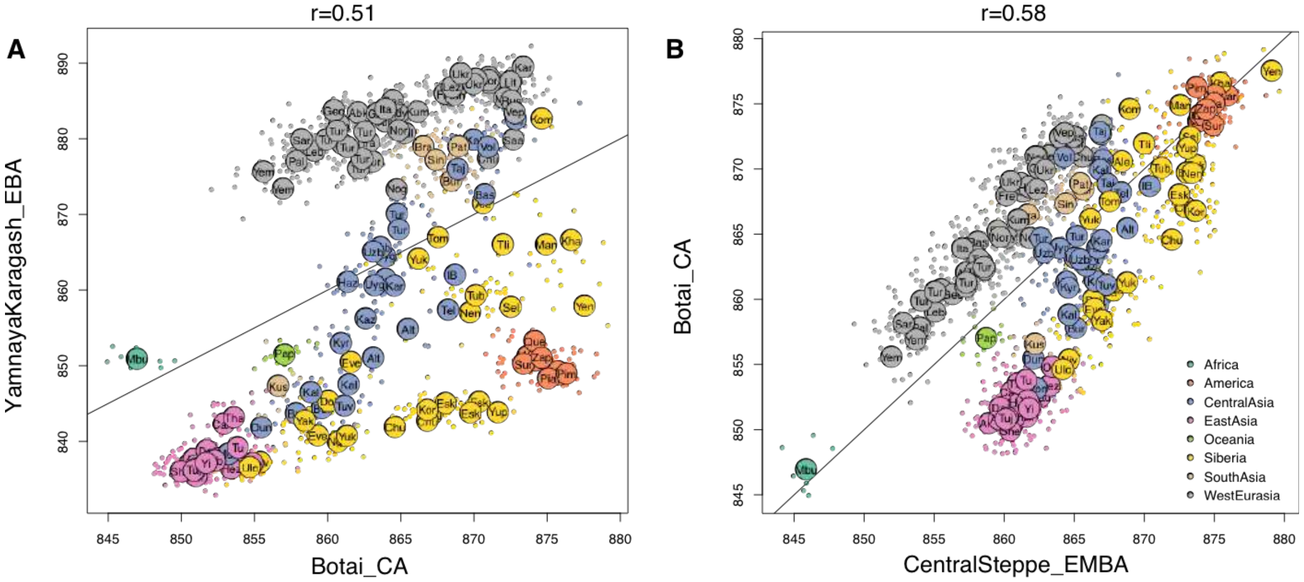


Fig. S31.

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).

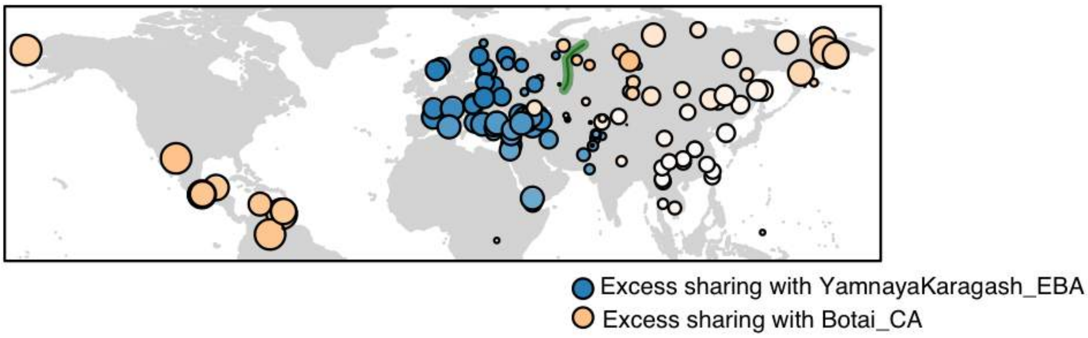


Fig. S32.

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.

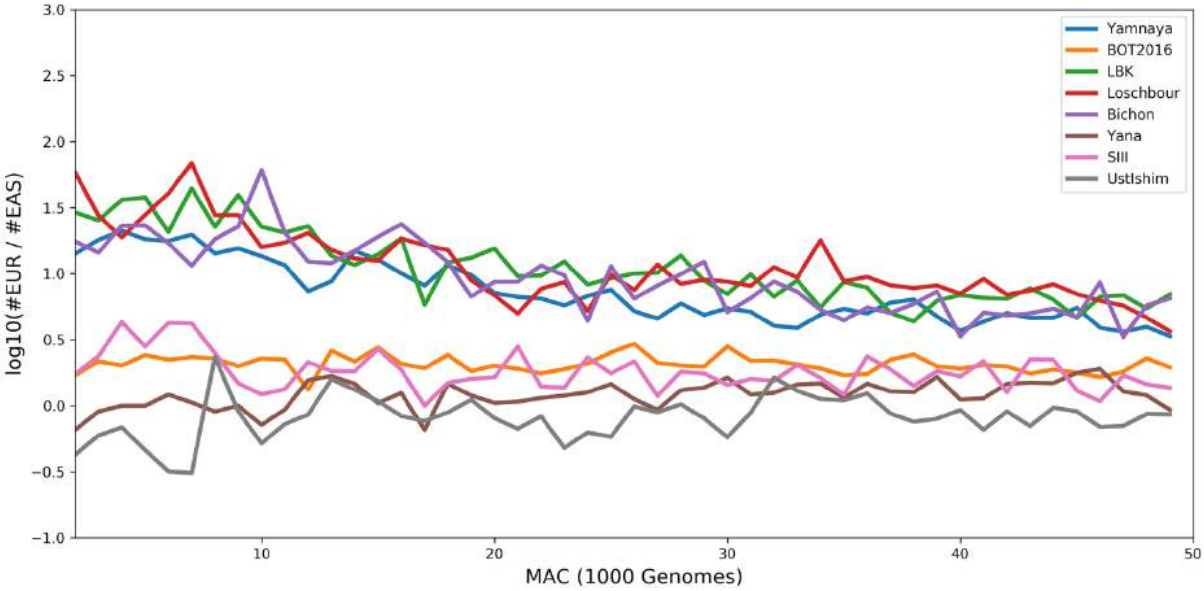


Fig. S33.

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).

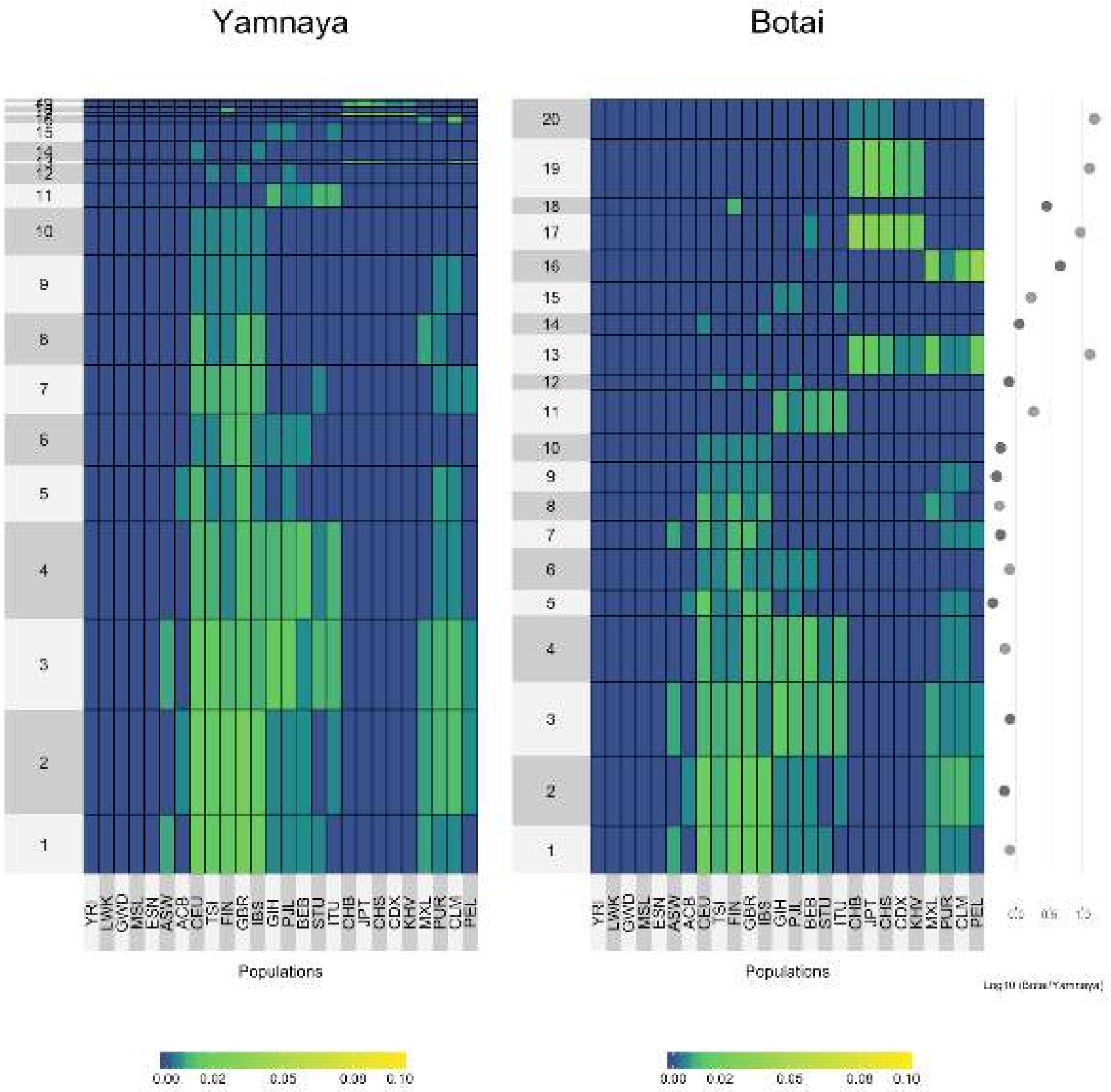


Fig. S34.

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.

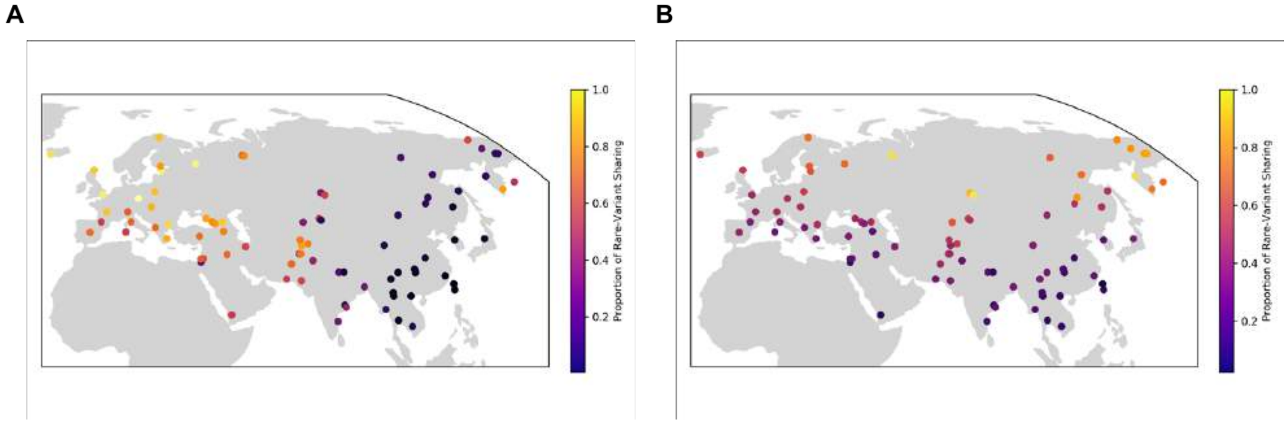


Fig. S35.

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

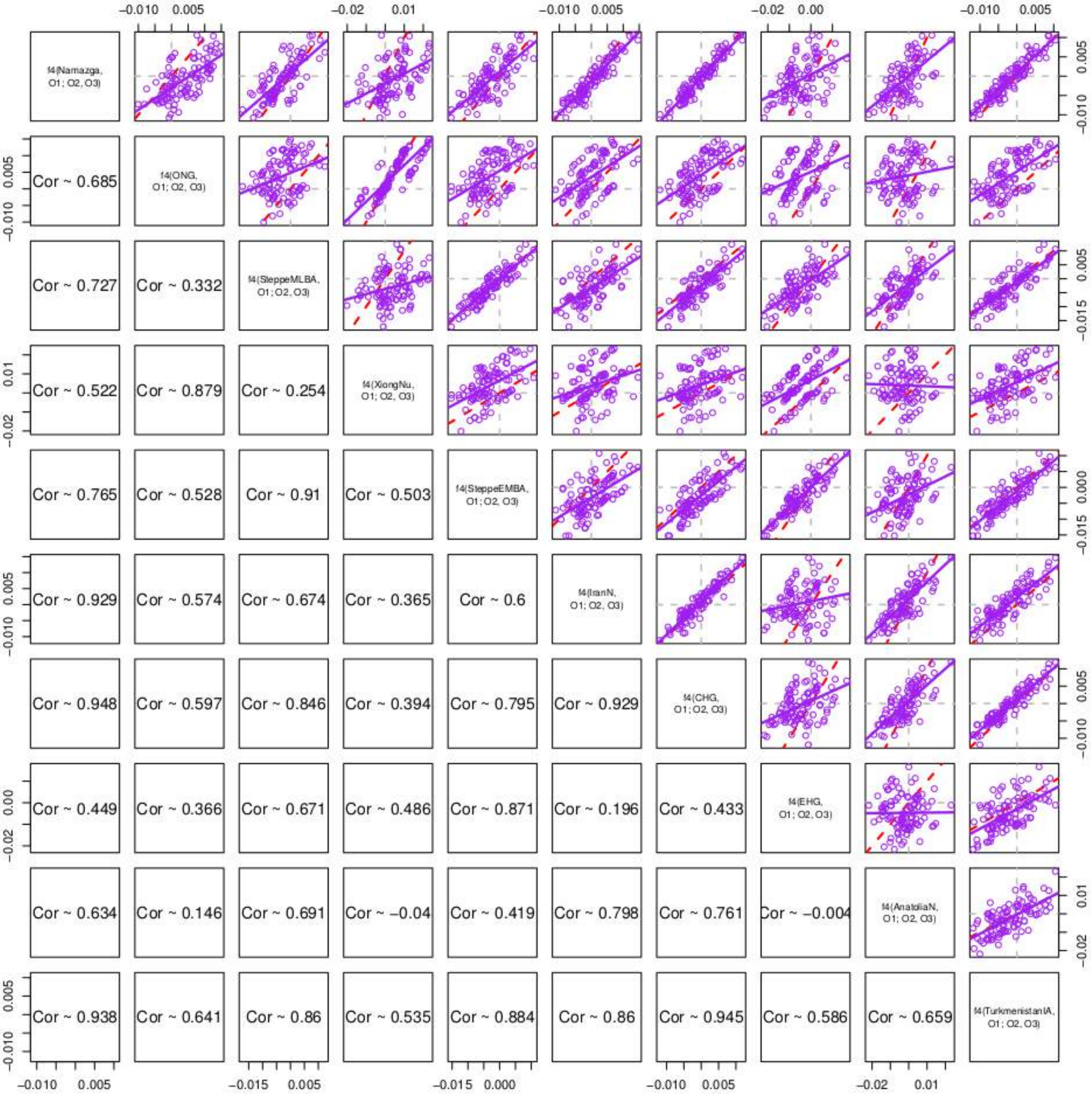


Fig. S36.

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).

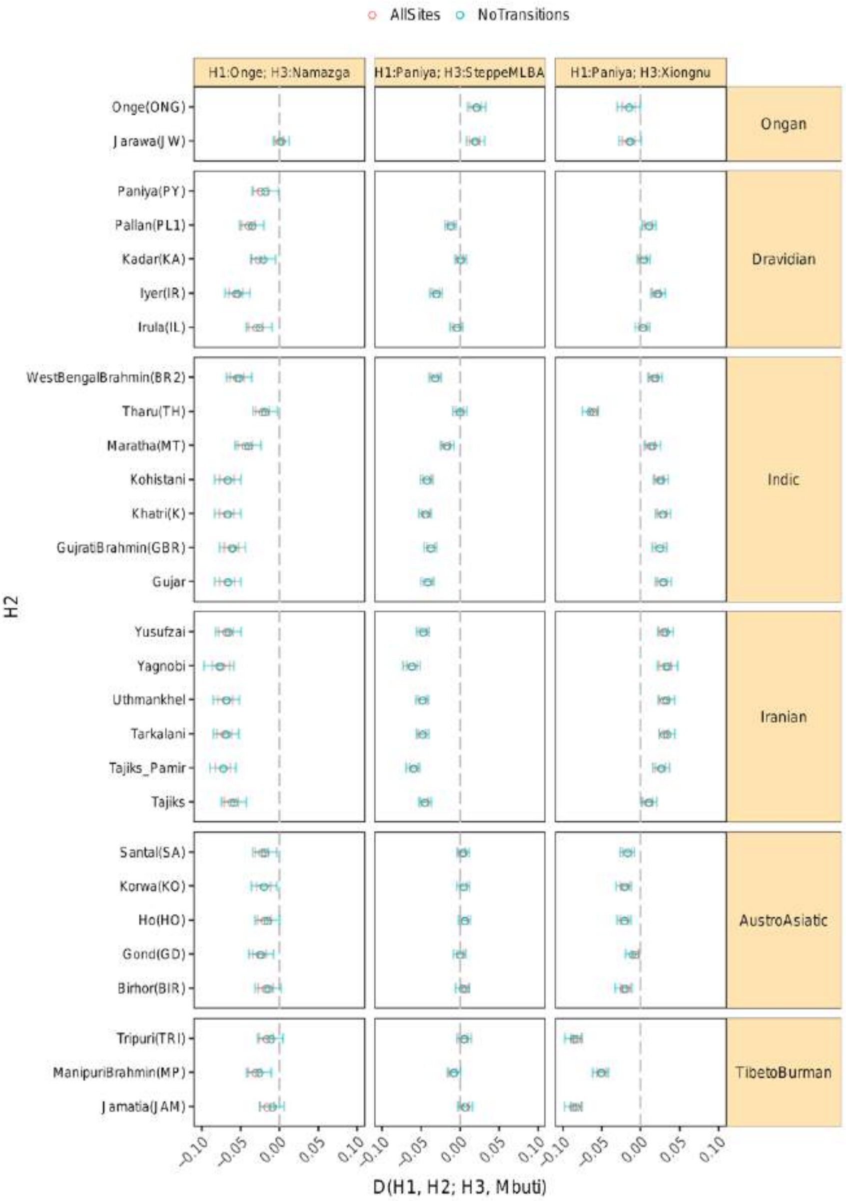


Fig. S37.

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
$N_{\mathrm{Ancestral}}$	12464.47	(12412.77, 12506.72)	-1.38	24.13
N_{Eurasia}	3683.89	(3616.90, 3777.24)	10.45	45.70
$T_{\mathrm{Mbuti\text{-}Eurasia}}$	122513.04	(121630.09, 123547.00)	34.46	513.67
$T_{\mathrm{AEA\text{-}ANE}}$	48294.75	(46577.94, 49632.08)	-129.92	754.35
$T_{\mathrm{Han\text{-}ShamankaEN}}$	17486.66	(16652.49, 18268.40)	-40.27	390.92
$T_{\mathrm{MA1\text{-}GhostANE}}$	26580.47	(25319.17, 29126.32)	591.97	896.66
$T_{\mathrm{Steppe\text{-}GhostANE}}$	12536.92	(11607.70, 19220.01)	2174.32	1874.60
$T_{\mathrm{AEA\text{-}>Steppe}}$	12535.30	(10924.24, 13976.98)	-119.99	813.02
$P_{\mathrm{AEA\text{-}>Steppe}}$	0.51	(0.49, 0.53)	0.00	0.01
$T_{\mathrm{Sholpan\text{-}Okunevo}}$	10545.15	(9528.28, 11378.47)	-101.47	490.01
T_{Botai}	11358.50	(8893.71, 12411.86)	-735.59	905.34
$P_{\mathrm{GhostANE\text{-}>Botai}}$	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

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

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.

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.

Archaeological supplement A to Damgaard et al. 2018: Archaeology of the Caucasus, Anatolia, Central and South Asia 4000-1500 BCE

AUTHORS

Kristian Kristiansen[1], Brian Hemphill[2], Gojko Barjamovic[3], Sachihiro Omura[4], Süleyman Yücel Senyurt[5], Vyacheslav Moiseyev[6], Andrey Gromov[6], Fulya Eylem Yediay[7], Habib Ahmad[8,9], Abdul Hameed[10], Abdul Samad[11], Nazish Gul[8], Muhammad Hassan Khokhar[12], and Peter de Barros Damgaard[13].

AFFILIATIONS

[1]Department of Historical Studies, University of Gothenburg, 40530 Göteborg, Sweden.
[2]Department of Anthropology, University of Alaska, Fairbanks, USA.
[3]Department of Near Eastern Languages and Civilizations, Harvard University, USA.
[4]Japanese Institute of Anatolian Archaeology, Kaman, Kırşehir, Turkey.
[5]Department of Archaeology, Faculty of Arts, Gazi University, Ankara, Turkey.
[6]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, Russia.
[7]The Institute of Forensic Sciences, Istanbul University, Istanbul, Turkey.
[8]Department of Genetics, Hazara University, Garden Campus, Mansehra, Pakistan.
[9]Islamia University, Peshawar, Pakistan.
[10]Department of Archeology, Hazara University, Garden Campus, Mansehra, Pakistan.
[11]Directorate of Archaeology and Museums Government of Khyber Pakhtunkhwa, Pakistan.
[12]Archaeological Museum Harappa at Archaeology Department Govt. of Punjab, Pakistan.
[13]Centre for GeoGenetics, Natural History Museum, University of Copenhagen.

ABSTRACT

We present a brief archaeological summary of the main phases of cultural and social change in the Western, Central, and South Asia ca. 4000-1500 BCE as a contextual framework for the findings presented in Damgaard et al. 2018. We stress the role of the Caucasus as a conduit in Western Asia linking the steppe and Eastern Europe with Anatolia, Syria, Iraq, and Iran. We track the emergence of the Bactria-Margiana Archaeological Complex (BMAC) in Central Asia as a cultural melting pot between the steppe and the sown lands during a period of more than a millennium. And we highlight indicators of cultural and commercial exchange, tracking developments in technology as well as social and political organization that came about as part of complex processes of interaction in a region stretching from South Asia to the Mediterranean.

1. Anatolia and Caucasus

We present a brief summary of the main phases of cultural and social change in the Caucasus and Anatolia from 4000–1500 BCE. Both were areas of dynamic mediation and innovation due to their control of rich mountain resources and their position between the steppe in the north and the urban civilizations of the south (Kohl 2007: ch. 3; Smith 2015; Wilkinson 2014).

1.1. 4th millennium BCE: innovations in textile production and metallurgy, expansion of trade, rise in mobility. The Uruk Expansion, and Maykop culture of the northern Caucasus.

During the 4th millennium BCE the Caucasus and Anatolia entered a period of dynamic exchange that coincides partially with the so-called “Uruk Expansion” in southern Mesopotamia (Sagona 2011). The latter is defined by an explosive growth in population, the rise of statehood, urbanization, technologies of communication, and a complete restructuring of social, political and commercial institutions (Algaze 1995). Surrounding the Mesopotamian urban centers along the mountainous arch that stretches from southwestern Iran to southern Turkey was a series of smaller settlements that shared their material and visual culture as well as their political institutions with the main cities. They seem to constitute a network of early trading posts that provided raw materials (timber, stone, metal, and possibly also workers) to the urban south, probably in return for costly textiles (Wilkinson in press: Fig. 3). Sites like Arslantepe on the Upper Euphrates in Turkey acted as conduits for this network and ultimately connected the dense urban regions to production sites as far away as the Caucasus.

Examples of southern luxury fabrics have been found in Maykop burials in the northern Caucasus (Kohl 2007: 72–86) together with an array of copper, gold, and silver objects, weapons, tools, buckets, and drinking cups. The Maykop culture of northern, and even southern Caucasus (Lyonnet et al. 2008), spread innovations in metallurgy and metalwork onto the steppe and eastern Europe (Hansen 2010), into Iran (Ivanova 2012), and into central Anatolia (Rahmstorf 2010: Fig. 3) as part of a cultural “bundle” that also included wheels, wagons, and knowledge of mining (Hansen 2014: Fig. 1). Along with the spread of goods and technology, we must assume that also people moved as traders and craftsmen in search of new sources of metal, patrons, and wealth. The expansion dates mostly to the late 4th millennium BCE, when we also find, e.g. at Arslantepe, a royal burial with connections to the northern Caucasus and the late Maykop culture, possibly as a sign of incoming new elites (Palumbi 2007) or dynastic intermarriage.

1.2. 3rd millennium BCE: Kura-Araxes semi-urban culture of Transcaucasia, eastern Anatolia and northern Mesopotamia, followed by Trialeti kurgan culture from 2100 BCE.

A second cultural group to emerge out of contact with the Uruk networks in Transcaucasia towards the end of the 4th millennium BCE was the Kura-Araxes / Early Transcaucasian Culture (ETC) (Kohl 2007: 86–102; Wilkinson 2014: 309–314). This "cultural historical community" remains poorly understood. It had a developed metal technology and fine pottery but shows little sign of social hierarchy. Most settlements are relatively small (under 5 ha), and the economy seems to have been mainly agrarian. The material culture is fairly homogenous across a large region in the Caucasus and Eastern Turkey with distinct assemblages stretching into Syria, the Levant, and western Iran. Its expansion has been associated with a sharp break at several central settlements of the former Uruk network as well as the introduction of new forms of architecture and material culture, again suggesting a movement of people.

Batiuk 2013 and Rothman 2015 have argued for a "rippled" process of migration from east to west, in which "push" factors in Transcaucasia and eastern Anatolia were balanced by "pull" factors in the destination zones. Batiuk 2013 used multiple lines of evidence, including settlement patterns, ceramic assemblages, and textual records, to postulate an association between the spread of ETC and the practice of viticulture, which has a long-recorded history in Transcaucasia. He states that the production of a consumable high-status commodity like wine by settlers moving west and identified by a use of Early Transcaucasian wares will have allowed these to keep a socioeconomic status and maintain a social identity in an archaeologically visible manner in their new homelands for extended periods of time.

It has been speculated whether settlers from the east also brought with them languages, such as early Hurrian or IE Anatolian. The personal names borne by individuals coming from the state of Armi in southern Anatolia attested in the archives as early as the 25th century BCE at Ebla (Archi 2011; Bonechi 1990) constitute a mixture of Semitic, Anatolian IE, and unknown background (Kroonen et al. 2018). A possible interpretation is that multiple groups moved into Anatolia from the Caucasus during the late 4th and early 3rd millennia BCE, including groups of proto-Hurrian and early IE Anatolian speakers. Clear from the written record of Bronze Age Anatolia, however, is also that language was not considered an ethnic marker there and that the region is characterized by its high population mobility and plurality of languages and traditions.

1.3. 2nd millennium BCE (2100–1500 BCE). The Trialeti royal kurgans, micro-polities, Old Assyrian traders, and the formation of the Hittite state.

By the end of the 3rd millennium, the Kura-Araxes, and Early Transcaucasian cultural sequence was broken by intrusions from the Caucasus, and ultimately from the steppe, seemingly associated with the re-emergence of royal kurgan mounds in Transcaucasia (Smith 2015: ch. 4) and a material horizon known as the Trialeti culture (Kohl 2007: 113–121). The kurgans, and with them a new subsistence economy based on herding, had already begun to spread towards Transcaucasia from the middle of the 3rd millennium BCE onwards. The movement reached its apex in the large and immensely rich kurgans characteristic of the Middle Bronze Age Trialeti. The mounds were constructed over huge timber-built burial chambers and had long stone paved procession roads leading to them (Narimanishvili and Shanshashvili 2010). These appear to be contemporaneous with the arrival of chariot warfare from the steppe (Kristiansen and Larsson 2005: Figure 79), and from the rich grave inventories it is clear that Trialeti elites traded with both Anatolia and northwestern Iran (Rubinson 2003). What they had to offer in return was probably silver and horses or mules, which begin to appear in Iran, Anatolia, and the Near East (Anthony 2007: 412–418, Fig. 16.3; Michel 2004). In return, they received prestige goods, such as golden drinking cups and fine textiles. There are cultural connections between Trialeti and the early Mycenaean shaft grave burials, presumably moving either via the steppe corridor or through Anatolia. The source of rich goods deposited in burials at both Mycenae and Trialeti appear to have come from Anatolia (Puturidze 2016).

During the Middle Bronze Age (ca. 2000–1650 BCE) Anatolia was divided into a number of micro-polities, probably numbering in the several hundred. Each was centered on an urban settlement and linked together in competitive and constantly shifting networks of political alliances that shared a common cultural and cultic horizon. Their history is reflected in the extensive archives kept by Old Assyrian merchants who operated a network of some forty trade settlements in Central Anatolia during the period in question (Barjamovic in press; Larsen 2015). They brought in tin and luxury textiles from distant Mesopotamia in return for silver and gold. Some 23,000 texts written on clay tablets in cuneiform signs reveal Anatolia as a multi-ethnic, polyglot, and cosmopolitan society with no visible markers (or even no clear notion of) any ethnic distinctions within the region. Instead, material and spiritual traditions were continually evolving into new and hybrid forms (Larsen and Lassen 2014) in a pattern that persisted also during the subsequent centuries after 1650 BCE under the centralized political authority of the emerging Hittite state. A polyglot, highly mobile, and culturally hybrid population renders a discussion of ethnic distinctions along linguistic lines meaningless, and rather, the situation seems to mirror historical and contemporary cases in which language is tied to function and not identity. Sources suggest that a given individual would speak a handful of languages, including perhaps one or two at home, a third in trade, and a fourth during cultic services, etc. Currently, there is evidence of Hattian, Hittite, Hurrian, Luwian, Akkadian (Assyrian/Babylonian), and Palaic speakers within Central Anatolia, with additional languages leaving little trace behind, except perhaps through personal names.

To conclude, we observe a changing dynamic between southern Mesopotamian and Caucasian influences into Anatolia and northwestern Iran between the 4th–2nd millennia BCE. The Caucasus served as a conduit linking the steppe and Eastern Europe with Anatolia and Iran as well as ultimately Mesopotamia and the Eastern Mediterranean. Influences from the Caucasus first reached Anatolia during the mid- to late 4th millennium, through the Maykop culture, which also influenced the formation and apparent westward migration of the Yamnaya. A second wave of steppe influences entered during the late 3rd and early 2nd millennia with the chariot horizon and the Trialeti culture. Both of these expansions had a steppe corridor route and an inland Anatolian route reaching the Aegean.

2. Central and South Asia

The following provides a summary of cultural developments observed in the archaeological record of populations residing within and adjacent to the piedmont strip located along the foothills of the Kopet Dagh mountains of southern Turkmenistan from the beginning of the Eneolithic Namazga culture (ca. 4000 BCE) to the end of the Bactria-Margiana Archaeological Complex (BMAC) during the middle of the 2nd millennium BCE. This is followed by a brief account of the Indus and Gandharan Cultures.

2.1. The Middle Eneolithic: Namazga [NMG] II (ca. 4000–3500 BCE)

The Middle Eneolithic was a time of considerable transformation. The Geoksyur oasis sites represent the easternmost sedentary agriculturalist communities whose neighbors would have been Neolithic hunting groups of the Kelteminar culture (Dolukhanov 1986). These ranged the nearby steppe and semi-desert regions further north and east. Moving northeastwards, it appears that part of the Geoksyur oasis population entered the northern reaches of the Murghab River delta, where a recent find exposed several widely scattered settlements with ceramics typical of the Geoksyur style (Salvatori 2008: 76).

During the Middle Eneolithic, lapis lazuli first came into systematic use. Efforts to provide a regular supply of this stone, whose main deposits lie in the mountains of northeastern Afghanistan (Badakhshan), likely played a significant role in the establishment of lasting trade and cultural ties over a vast territory. To Salvatori 2008: 76), the long-distance contacts of the Geoksyurian population to the east at Sarazm and perhaps to the northern reaches of the Murghab delta closer by at Kelleli 1 laid the exploratory foundation for considerable and extensive geographic knowledge as well as for an outward worldview in the quest for highly prized and non-locally available resources—a quest that only intensified over time and that characterizes the Late Eneolithic and Bronze Age in southern Turkmenistan. Yet, despite the far-flung contacts to the south, southeast, and east for the acquisition of metallic ores and semiprecious stones, there is no evidence of contact across the Aral Basin with the Neolithic populations of the steppe zone to the north (Hiebert 2002).

2.2. The Late Eneolithic: Namazga [NMG] III (ca. 3500–3000 BCE)

The Late Eneolithic period is marked by a general continuation of the trends observable during the previous period. Throughout southern Turkmenistan there is a tendency for the major sites in a particular area to increase in size and for the overall number of sites within the region to decrease. Settlements appear to have been pre-planned and the multi-chambered residential units with their own courtyard characteristic also of the NMG II period continue into the Late Eneolithic (Masson 1992: 231). While the archaeological record provides abundant evidence for an array of contacts between populations of southern Turkmenistan and populations to the east (Sarazm) and south (Baluchistan, Seistan) during the Late Eneolithic, there is no evidence for any substantial contacts between NMG III populations to populations occupying the steppe zone to the north.

2.3. Early Bronze or “Proto-Urban” Period [NMG IV] (ca. 3000–2500 BCE)

The Early Bronze or “proto-urban” period appears to have been an age of important technological and social development but is less well-understood than the preceding Late Eneolithic and subsequent Middle Bronze periods. Technological developments include the introduction and increasing use of the potter’s wheel, improved furnaces for smelting copper, and the beginnings of monumental architecture. It also features a separation of settlements into either large, proto-urban sites (*e.g.*, Namazga-depe, Ulug-depe, Khapuz-depe, and Altyn-depe) or small villages.

2.4. Middle Bronze Age, NMG V, and the BMAC culture 2500–2000/1900 BCE

The term BMAC (Bactria-Margiana Archaeological Complex) is commonly used for the phase after 2500 BCE and also is sometimes called the Oxus civilization. The processes behind the formation, florescence, and dissolution of the BMAC culture remains poorly understood. Around 300 settlements are known, many of them heavily fortified. There is a rich material culture with links to steppe cultures, the Indus, and Iranian cultures (Kohl 2007: ch. 5; Parpola 2015: ch. 8).

Some have argued that the final BMAC is a candidate for one of the expansions of Indo-Iranian language to northern India/Pakistan and the Iranian plateau (Parpola 2015). Others would see the chariot riding pastoralists of the Sintashta and later Andronovo cultural horizons as the original cultural and linguistic influence behind Indo-Iranian (Kuzmina 2006). We return to the complexity of the situation below and presently address only the cultural and archaeological sequences and the interaction of BMAC with steppe, northern India/Pakistan, and Iran as it represented a cultural, and probably also a genetic and linguistic melting pot between the steppe and south Asia. We present a brief summary of the main cultural phases followed by a discussion of interactions as reflected in the archaeological record.

The Middle Bronze Age (2500–1900 BCE) represents a developed urban civilization based on irrigation but with a large settlement area stretching outside the oases. New standards appear in nearly all aspects of culture. A complex hierarchical settlement pattern suggests a developed political organization that ended around 1900 BCE with a collapse of the major settlements and a marked reduction in size when rebuilt (Salvatori 2016). New smaller settlements were constructed with fortified walls and round towers, which suggest smaller political units. This change in settlement pattern has been linked by some to the first arrival of steppe metal objects and pottery (Anthony 2007: Fig. 16.6) and hybrid burials that combine BMAC and steppe grave goods (Anthony 2007: Fig. 16.8; Kohl 2007: 208–209). Central Asian trade goods also appear in the steppe (Anthony 2007: 433–434).

2.5. The Late Bronze Age (1900–1750 BCE) and Final Bronze Age (1750–1500 BCE).

Understanding the gradual decline and final disintegration of the oasis civilization towards the end of the Bronze Age continues to defy common consensus (Kohl 2007: ch. 5). Increasing numbers of steppe pastoralists probably moved south and settled around the oases, but one could argue that trade with steppe populations was a driver behind some of these changes. At the time, Andronovo groups seemingly controlled the tin production and distribution from mines in central Asia (Parzinger 2003), which may help to explain their increasing influence and expansion into areas occupied BMAC populations. Their presence is reflected in numerous campsites and may have been a contributing factor in the final collapse of the BMAC settlements around 1500 BCE (Spengler et al. 2014).

2.6. Concluding Remarks

Located at the crossroads between different environmental and cultural zones and bounded by the Caspian Sea to the west, the steppe and steppe/desert of Kara Kum to the north, the Iranian plateau to the south and southwest, and the Indus cities to the southeast, the Bactrian-Margianan Archaeological Complex emerged as a cultural melting pot between the steppe and the sown lands during a period of more than a millennium. It formed a distinct social and cultural entity located along a fertile strip of land just 80 km wide and 600 km long. It flourished during an arid period from 2400 BCE onwards, making the fertile land attractive to newcomers from both north and south and lending to it a characteristic cultural, and probably also genetic and linguistic admixture.

2.7. Indus-Harappa Culture

The Indus Valley is largely located within present-day Pakistan and northern India, its watershed stretching from the Chinese frontier in the northeast, and bordering onto Afghanistan in the north and Iran in the west. The known settlement chronology of the area spans from the Neolithic Mehrgarh I–VI phases (7000–3300 BCE) to the Bronze Age, with its earliest evidence of the Indus Valley Civilization (IVC) coming from the Harappa site ca. 2800 BCE. Scholars have suggested that populations forming the early Harappan phases of the IVC were farmers and lived here in very large numbers, up to 1 million people. The Indus civilization reached its high point during the period 2600–1900 BCE, with an overall standardization of material culture and a four-tiered settlement hierarchy across an area of roughly 500,000 sq. km. Since its primary urban centers were located mainly in the lowland floodplain, the cities were in need of importing most of their raw materials, including metal, stone, and quality timber, from beyond its area of control. This led to an extensive trade network with outside regions, and Indus cities maintained trade with faraway partners in Afghanistan, Iran and BMAC, the Persian Gulf, and Mesopotamia (Laursen and Steinkeller 2017; Ratnagar 2006; Wright 2010).

The process and causes behind the decline of the Indus civilization after 1900 BCE are poorly understood, but they included an abandonment of the large urban settlements as well as the script and homogenous material culture associated with them. The region seemingly dissolved into smaller local and regional groups (Francfort 2001; Franke-Vogt 2001).

2.8. Gandhara Grave Culture

The Gandhara grave sites were initially reported by the Italian Archaeological Mission to Pakistan and the Department of Archaeology at the University of Peshawar. These graves were first reported from the Swat and Dir regions of ancient Gandhara, a region, which is said to have extended from the western boundary to the Peshawar Valley to the Indus in the east and comprised the hilly tracts south of the river Swat and Buner in the north (Hassan 2013: 3). It was this region, which gave birth the Buddhist civilization of Gandhara, that emerged during the 3rd century BCE under the Mauryans (Hassan 2013: 5) and later on flourished under the Indo Greeks, Scythians, Parthians, Kushans, and Sassanians up until the invasion of the White Huns in the 5th century CE, who are held responsible for the decline of this civilization (Marshall 1951: 285).

The Gandharan Grave culture predates the Buddhist civilization of Gandhara. The term Gandharan Grave culture was coined by Dani (1968: 99) after having discovered and excavated many grave sites in the Dir, Swat, and Bajaur regions of Pakistan. Later archaeological surveys and excavations conducted by both Pakistani archaeologists and foreign missions revealed many similar grave sites outside of the Gandharan region, indicating that this culture was not confined to ancient Gandhara but rather extended to include parts of the Chitral and Mansehra Districts of the present-day Khyber Pakhtunkhwa (KP) Province, Pakistan. Of these, the former district encompasses the greatest number of Gandharan Grave culture sites. Excavations at the grave sites in Chitral, especially at the sites of Parwak (Ali and Zahir 2005) in 2003–2004 and Singoor sites by the Directorate of Archaeology and Museums Government of KP in 2005, and later at Gankoreneotek (Ali et al. 2010) and Chakast sites by the Department of Archaeology at Hazara University, Mansehra, have yielded artifacts and skeletal remains. While recent archaeological surveys in the latter district have resulted in the discovery of four sites: Chansoor Dheri I, II and III as well as Naukot. Of these, the Chansoor Dheri III was accidently discovered during the construction by the owner. This site yielded an urn burial of a male adult and a terracotta bead (Figs. 1, 2, and 3) (Hameed 2012: 14–15). Recent research has emphasized the complexity of the Gandharan Grave culture as resulting from both local and external processes (Zahir 2016). Recently obtained C14 dates place the Gandhara grave culture between 1000 BCE and 1000 CE (Ali et al. 2008), with those from Chitral District being more recent from those found in the lowlands of the Dir and Swat Districts.

3. Interactions between the Settled Communities of Southern Central Asia and Steppe Populations during the Bronze Age and their Relationship to the Gandharan Grave Culture of Northwestern Pakistan

3.1. Middle Bronze Age, NMG VI, and the BMAC ca. 2500–1500 BCE

The transition from the Middle to the Late Bronze Age (MG VI) is a time of considerable change in southern Central Asia. It was once believed that the large urban centers of the Kopet Dagh piedmont suffered some kind of “urban crisis” near the end of the Middle Bronze Age (Biscione 1977; Hiebert 1994: 174–75, 2002b; Masson, 1992b: 342). However, it now appears that settlement of Margiana and perhaps the Bactrian oases occurred, not after the NMG V period, but contemporaneously with its later temporal range (ca. 2200–2000 BCE: Salvatori 2008: 77). It now seems more likely that the colonization of Margiana was, in fact, a consequence of population movement from the Kopet Dagh foothills, but rather than occurring at a time of crisis, it occurred when Altyn-depe was at its peak size (Hiebert 1994; Masson, 1992a). This is attested by the close similarities in ceramics, small finds, and architecture found in the deepest strata at numerous sites in Margiana to those found at contemporaneous NMG V deposits in the urban centers of the Kopet Dagh (Hiebert and Lamberg-Karlovsky 1992: 4; P’yankova 1989; Salvatori 1994; Sarianidi 1990; Udemuradov 1986). Whether this colonization from the piedmont extended further to encompass the Bactrian oases situated along tributaries of the Amu Daya to the east is the subject of considerable controversy (Francfort 1984; Hiebert 1994; Khlopina 1972: 213–14; Sarianidi 1999).

It has long been assumed, because of close correspondences in artifact assemblages, architecture, and inhumation practices, that populations of the Kopet Dagh piedmont urban centers first settled in Margiana through a process of segmentation and that a portion of this population subsequently moved further east to establish urban centers in the unpopulated northern, southern, and eastern Bactrian oases (Boroffka et al. 2002: 138; Hiebert 1994; Masson 1992b: 345). Francfort 1984 finds this scenario unlikely for several reasons. First, given populations known to be found in arable lands to the north (Zaman Baba culture of the mid- to lower Zarafshan Valley, Sarazm of the Zarafshan Valley) and east (Shortughaï, a Harappan outpost located in the eastern Bactrian oasis) it is unlikely the northern and southern Bactrian oases were unpopulated. Second, radiocarbon dates from the northern Bactrian urban center of Sapalli-Tepe are contemporaneous, not subsequent to the earliest settlements in Margiana (see Salvatori 2008). Third, there are numerous stylistic differences, especially with regard to the bronze pins and seals that distinguish small finds at Bactrian sites from those at sites in Margiana (Francfort 1984). Perhaps the most telling difference in the artifact assemblages from Bactria to those from Margiana involves the elemental composition of the bronze objects.

Metallurgical technology has a long history in southern Central Asia that likely can be traced to influences from Iran (Kohl 1984: 71). At sites in the Kopet Dagh piedmont and in Margiana, bronze objects are almost exclusively alloyed with lead and/or arsenic (Anthony 2007: 420; Gupta 1979; Hiebert 1994: 159–60; Hiebert and Killick 1993: 199; Masson and Kiiatkina 1981; Salvatori et al. 2003: 79; Terekhova 1981: 319). In contrast, the metal assemblages recovered from such BMAC sites in northern Bactria as Djarkutan and Sapalli-Tepe feature bronze that is alloyed with tin, which may account for as much as 50% of all bronze objects (Anthony 2007; Chernykh 1992: 176–82; Salvatori et al. 2003: 79). Hence, it appears that there were two centers of metallurgical production in southern Central Asia across the transition from the 3rd to the 2nd millennia BCE (Chernykh 1992: 179; Francfort 1984; Hiebert 1994: 384). Indeed, later ceramic assemblages from sites in Margiana (Hiebert 1994's Takhirbai phase) and the latest Bronze Age occupation of the Kopet Dagh piedmont (the so-called NMG VI) containing the deeply burnished gray wares characteristic of northern Bactria suggest that cultural influences likely flowed from east to west, rather than exclusively from west to east as has long been assumed (Francfort 1979, 1984, 1989; Kohl 1993; but see Heibert 1994: 68–69). This dynamic, when coupled with the probable presence of a local resident population within the Bactrian oases prior to the Middle Bronze Age, likely accounts for the fact that phenetic affinities between the Middle Bronze Age inhabitants of Altyn-depe and those of northern Bactria are not especially close (Hemphill 1999b, 2013; Hemphill and Mallory 2004).

3.2. Interactions between BMAC Populations and Steppe Bronze Populations

The Late Bronze Age in this region is known as the BMAC (Bactria-Margiana Archaeological Complex) in existence from ca. 2200-1500 BCE. The factors surrounding its formation, efflorescence, and dissolution remain enigmatic. Around 300 settlements are known, many of them marked by substantial fortifications. High quality wheel-thrown ceramic vessels were produced on an industrial scale and are found widely distributed throughout southwestern Central Asia (P'yankova 1989, 1994), including sites attributed to the Andronovo affiliated Tazabag'yab culture of the Aral Sea region (Khorezm) (Kohl 1993), the Zaman Baba culture of the middle Zarafshan Valley (Askarov 1962, 1981; Sarianidi 1979), as well as the Andronovo affiliated Vakhsh/Beskent cultures of southern Tajikistan (Kohl 1984, 1993).

BMAC artifacts have been discovered at a wide array of sites located on the Iranian Plateau as well as at the western margin of the Indus Valley (Hiebert 1994; Hiebert and Lamberg-Karlosky 1992; Jarrige 1994; Jarrige and Hassan 1989; Kohl 1993; Santoni 1984; Sarianidi 1999). These artifacts are not randomly present at these sites but tend to be associated with funerary contexts and include characteristic miniature columns of alabaster as well as bronze pins, brooches, and seals with characteristic BMAC motifs (Amiet 1986, 1989; Francfort 1994: 406–18; Hiebert 1994; Hiebert and Lamberg-Karlovsky 1992; Sarianidi 1981). Intriguingly, the presence of non-BMAC artifacts at BMAC sites in Bactria and Margiana are exceedingly rare (Hiebert 1994 164, 366; Hiebert and Lamberg-Karlovsky 1992: 12).

This unidirectional dynamic has led some researchers to consider the BMAC to have been a brief-lived imperial state (Hiebert and Lamberg-Karlovsky 1992: 12) while others see the BMAC as one of a number of participants in a vast *koiné* that involved populations residing on and about the peripheries of the Iranian Plateau (Anthony 2007; Jarrige 1994; Jarrige and Hassan 1989; Salvatori 1995; Salvatori et al. 2003; Santoni 1984). Coalescing during the mid-3rd millennium BCE and lasting to the end of the first quarter of the 2nd millennium BCE, these networks facilitated the circulation of highly desired “prestige” goods among elites, and among these were small finds made of tin bronze. It appears clear that the production center for the tin bronze objects was northern Bactria, but the origin of the tin-bearing remains debated.

Anthony 2007 has claimed recently that the discovery of tin mines along the Zarafshan River, the presence of a Petrovka settlement of Tugai 27 km west of Sarazm in the upper Zarafshan Valley, and a grave at Zardcha-Khalifa (1 km from Sarazm) all attest to: 1) the mining of tin by steppe bronze culture populations, 2) the presence of steppe populations in Khorezm near to BMAC populations in northern Bactria, and 3) actual contact between steppe bronze populations and BMAC. From this he constructs a scenario in which the southward expansion of these steppe populations wrested control over the trade in minerals and pastoral products, while their chariots gave them a military advantage over the oases and settlements of the BMAC resulting in the dissolution of this polity (Anthony 2007: 452–54). Each of these claims deserves close examination.

The tin mines include Mushiston, located 40 km east of Sarazm in the upper Zarafshan Valley, and Karnab in the middle Zarafshan Valley some 170 km west of Sarazm close to where sites of the Zaman Baba culture have been found. At the former, excavations of Boroffka et al. 2002: 141 revealed the presence of vast deposits of copper and tin in the form of stannite. However, it appears that the prehistoric miners who worked the deposits were not interested in the primary ore, for all of the ancient workings are in an oxidation zone containing secondary mineral, such as malachite that contain copper and cassiterite and others, which contain tin. Deep inside the excavated galleries were found several stone-grooved hammers and a few potsherds attributable to the Andonovo horizon. However, a wooden beam found in association with these artifacts yielded a radiocarbon date of 1515–1265 cal BCE, which almost completely postdates the BMAC.

Ancient mining activity in the middle Zarafshan Valley was initially identified by Litvinsky 1962 in the 1940s–1950s at Karnab and Changali. Karnab was reinvestigated by Boroffka et al. 2002: 145, who found the cassiterite ore to be very low in tin with concentrations usually less than 3%. During excavations at the site 20 stone hammers and additional stone tool fragments were recovered, along with sherds of typical Andronovo horizon ceramics. A radiocarbon date was not obtainable from the strata in which these artifacts were recovered, but a date from the stratum above it yielded a date between 905–705 cal BCE. Thus, there is evidence for a steppe presence and the mining of tin, but there is no evidence so far that this tin mining was contemporaneous with the BMAC.

Excavation at the site of Tugai revealed the presence of copper-smelting furnaces, crucibles with copper slag still adhering to them, a bronze celt, and the remains of a semi-subterranean house (Kuzmina 2001: 20–21). Ceramic vessels were recovered and these have been identified by the excavator (Avanesova 1996) as attributable to the Petrovka culture, an eastern offshoot of the Sintashta complex. Recent revision of the chronology of the various steppe archaeological cultures by Hanks et al. 2007: 362, Fig. 4 places the Petrovka culture between 1950–1675 BCE, which overlaps considerably with the BMAC. Indeed, this contemporaneity is attested by the recovery of several red polished ware vessels that Kuzmina 2001: 21 finds similar to those found in Baluchistan and the Indus Valley as well as a black burnished vessel whose closest parallels are to be found in the BMAC assemblages of northern Bactria. Kuzmina attributes the presence of these vessels at Tugai to contacts with the inhabitants of Sarazm. If so, such contacts reaffirm contacts between the urban centers of the northern Bactrian oasis (*i.e.*, SapalliTepe, possibly Djarkutan) and Sarazm. Continued smelting of copper without alloying with tin to produce bronze at Tugai suggests that alloying technology had not yet reached populations in this region of Central Asia, a curious finding given the presence of bronze to the north (Sintashta) and to the south (BMAC) at this time and not at all expected of a new hegemonic presence. This may indicate that the source of the tin found in the bronze artifacts in northern Bactrian BMAC assemblages came from somewhere else.

The much-discussed burial at Zardcha-Khalifa (Anthony 2007: Fig. 16.8; Kohl 2007: Fig. 5.15), located in Pendzhikent along the left bank of the Zarafshan River, consists of an oval grave within which are the remains of a male buried on his right side in a flexed position with his head to the southwest (Bobomulleov 1997; Bostonguhar 1998). The right arm was placed under his head while the left was positioned on his stomach. The remains of a horned ram are at his head. Such funerary treatment is typical of the Bactrian BMAC (Askarov 1977, 1981). The deceased is accompanied by a wealth of grave goods, and these include fine-quality wheel-thrown pink-colored globular vessels with narrow necks identical to those associated with the Djarkutan phase of the BMAC (Abdullaev 1979; Askarov and Abdullaev 1983). Of special interest is a bronze pin some 18 cm long topped with the figure of a horse (Kuzmina 2001: 23, Fig. 4.3). Pins with zoomorphic heads are widely known from Bactrian BMAC burials, but none of these show depictions of horses (Kuzmina 2001: 24). Anthony 2007: 431 interprets this grave as that of an immigrant from the north who had acquired many BMAC luxury goods. However, it is equally likely that this individual may have been a resident of one of the Bactrian urban centers of the BMAC who married into or traded with the local population residing along the upper Zarafshan Valley.

Tugai represents just one of a whole series of sites found across a wide swath of Central Asia that have been designated as “steppe bronze cultures” (Masson 1992b). The hand-made ceramic wares recovered from these sites are commonly attributed to the Alakul and Federovo variants of the Andronovo horizon, which have been radiocarbon dated to the period between 1900–1500 cal. BCE (Hanks et al. 2007: 362). The economy of these groups appears to have been a highly variable combination of animal husbandry and cultivation (Lightfoot et al. 2015), with cattle predominating among the livestock. Masson 1992a: 243 maintains that contacts with the sedentary populations of the BMAC provided a stimulus that resulted in economic changes and population growth during the first half of the 2nd millennium BCE as reflected by a dramatic increase in the number of steppe bronze sites. Indeed, Masson suggests that the southward expansion of these steppe-derived populations was met by an equal northward expansion of sedentary farming populations of the BMAC leading to greater sedentism and a greater reliance on agriculture among members of these steppe bronze cultures. Kohl 2002: 78 agrees, arguing that these cow herders from the north changed their way of life and material culture when they entered this more developed sedentary world

Two good examples of this cultural hybridization process are the Tazabag’yab culture (Tolstov 1962; Tolstov et al. 1963), which is known from some 50 sites located within the Amu Darya delta, and the Zaman Baba culture (Askarov 1962, 1981; Gulyamov et al. 1966; Sarianidi 1979), which is represented by sites along the lower reaches of the Zarafshan River (Gupta 1979; Masson 1992a). At these localities, populations resided in sedentary villages, raised crops of wheat and barley on irrigated fields, raised domesticated cows, sheep, and goats, used hand-thrown steppe ceramics, but utilized bronze objects whose closest parallels are with those recovered from Bactrian BMAC sites (Masson 1992a, Kohl 1992) However, they employed catacomb burials of steppe type as well (Alekshin 1986: 92; but see Khlopin 1989: 83, 1994: 364–366). Thus, a complex cultural interaction whose exact nature is still debated took place between BMAC farming communities to the south and the Eurasian steppe societies.

Much has been made of the fact that ceramic wares attributable to steppe bronze cultures have been found at such BMAC sites (Anthony 2007: 427–33; Heibert 1994: 69; Kuzmina 1986, 2003; Lamberg-Karlovsky 2002; P’yankova 1993: 116; Vinogradova and Kuzmina, 1986) and has led Lamberg-Karlovsky 2002 to conclude that there is little doubt that the nature of interaction between steppe and settled BMAC populations was both extensive and intensive, if not always peaceful. Yet, the specific Andronovo horizon from which such sherds are attributed varies from site to site and by researcher (Hiebert 1994: 70; P’yankova 1993: 115–16; Vinogradova and Kuzmina 1986: Fig. 3), and these sherds appear to be more common at sites in northern Bactria than in Margiana (Hiebert 1994: 70).

Sarianidi 1998: 42, 1990: 63 has long been adamant that the steppe presence in Margiana and Bactria during the BMAC has been much overstated, noting that “pottery of the Andronovo type does not exceed 100 fragments in all of southern Turkmenistan.” Apart from the beheaded remains of a foal adjacent to the so-called “royal tomb” at Gonur North, excavations of BMAC sites have failed to yield horse remains among the animal bones recovered from these sites. The “royal tomb” itself, however, yielded grave goods that included a bronze image of a horse’s head on what may have been the pommel of a wooden staff. Another horse head image was found on a crested copper axe obtained on the art market, and a BMAC-style seal, likely looted from a cemetery in southern Bactria, depicts a man riding atop a galloping equine that looks like a horse (see Anthony 2007: 427). Such evidence suggests the BMAC inhabitants of the Bactrian oases knew about horses but did not eat them or apparently place much interest in them.

A similar situation exists for other aspects of steppe culture often associated with the presence of Indo-Aryan or proto-Indo-Aryan speakers. Sarianidi 1981, 1993, 1999 has proposed that the BMAC be considered as an intrusion of Indo-Aryans based upon two lines of evidence. The first is the presence of a possible elite social stratum due to the recovery of ritual axes with horse head motifs. The second is the presence of Andronovo-style ceramic wares located in special “white rooms” used for the preparation of a ritual drink (*haoma* in the Iranian *Avesta*, *soma* in the Indic *Rig Veda*). Parpola 1988, 1993, 1995 has taken Sarianidi’s thesis further, not only suggesting that the BMAC urban centers signal the adoption of a new strongly stratified social system evidenced by luxury goods, monolithic architecture, fortifications as well as the construction and maintenance of complex irrigation works, but also suggesting that these northern invaders came in two waves: the first were the proto-Aryans (Dasas) during the early BMAC, while the second were the Aryans of the *Rig Veda* during the late BMAC as witnessed by the “white rooms” at Gonur South in Margiana and at Togolok 21 in Bactria.

Subsequent research, however, has failed to support the claims of Sarianidi and Parpola. As noted by Francfort 1992, there is nothing in the rich iconography of the BMAC that presents features that could be considered Proto-Indo-Aryan or Indo-Aryan. Examination of the seed impressions from vessels found in the “white rooms” at Gonur South and Togolok 21, claimed to contain impressions of the *Cannabis* and *Ephedra* used to make the ritual drink, were identified by palaeobotanists at Helsinki and Leiden University as likely made by broomcorn millet (*Panicum miliaceum*) (Bakkels 2003).

According to Lamberg-Karlovsky 2002: 71, Margiana, Bactria, and adjacent lands to the north (Khorezm) were what Pratt 1992: 6–7 calls a contact zone “in which peoples geographically and historically separated come into contact with each other” characterized by “radically asymmetrical relations of power.” (p. 71). Yet, while the movement of steppe influences far to the south, extending up to the middle reaches of the Amu Darya, is indisputable (Masson 1992b: 335–36), there are no traces of a violent incursion by warlike steppe-dwellers into the ancient cities (Lyonnet 1994). There is no evidence of burning, no evidence of systematic destruction, and apart from an alleged “sacrificial” tomb at Gonur South (Sarianidi 2008), no evidence of violent deaths.

The archaeological record shows an interaction between the world of the steppes and the settled agriculturalists on the plains of Bactria and Margiana (Anthony 2007: ch. 16). That record also documents a process of assimilation between peoples from the north with sedentary agriculturalists who already participated in a greater cultural tradition with millennia-old roots extending back into southern Turkmenistan and Baluchistan (Salvatori 2008). This is further supported by anthropological analyses (Hemphill and Mallory 2004). However, waters divide when it comes to the interpretation of the nature and implications of these interactions between the steppe and the sown.

3.3. Vakhsk/Beshkent Cultures and the Gandharan Grave Culture

Alekshin 1986 maintains that not all of the steppe people who came to southwestern Central Asia in the early 2nd millennium BCE became farmers, some leading to the formation of the Vakhsh and Beshkent cultures found in the valleys of southern Tajikistan. Here, settlement sites such as Kangurt-tut and Teguzak (Kohl 1992: 192; Negmatov 1982: 61; Vinogradova 1993: 292, 294) have ceramic parallels with the northern Bactrian BMAC wares of the Molali phase, while cemetery sites, such as the Vakhsh catacomb burials, which are noteworthy for their elaborate construction with *dromoi* entrances and ritual use of fire, also contain metal artifacts that are similar to objects found on the northern steppes (Francfort 1981; Kohl 1992; P’yankova 1994: 369).

Parpola 1995 has suggested that the Vakhsh/Beshkent cultures correspond temporally with the Molali phase and are associated with the collapse of the BMAC. He writes, “It seems conceivable that nomadic tribes associated with the Vakhsh and Bishkent cultures took over the BMAC, as was once argued (Biscione 1977; Parpola 1988). Yet, there is no visible ‘Andronovoisation’ of the culture ... This suggests that, once again, the conquerors had quickly taken over, and adapted themselves to, the earlier local culture, the BMAC" (Parpola 1995: 10). Thus, in Parpola’s view, the BMAC was taken over by these semi-sedentary steppe nomads living in the adjacent highlands to the northeast of Bactria through a fairly peaceful *coup d’état*. This seems debatable given the low number and small size of sites attributed to these cultures. However, it illustrates the difficulty of interpreting political and social dominance from the archaeological record.

An alternative view has been offered by Vinogradova 1993: 300, who suggests that Vakhsh/Beshkent populations served as traders in a north-south exchange system along the western margin of the Pamirs. In her view, they served as the southern contact obtaining agricultural produce and ceramic wares from BMAC populations and moving these commodities northward in exchange with their northern counterparts in the upper Zarafshan Valley near Sarazm for tin and other metal ores. It may be that this trading conduit extended even further to the south to Shortughaï and the eastern Bactrian oasis of northeastern Afghanistan and beyond (Vinogradova 1993: 300).

Drawing parallels between the Vakhsh/Beshkent cemeteries and those of the Gandharan Grave culture, Parpola 1995 has proposed that such connections may have spanned the Hindu Kush and spread to the valleys of Dir and Swat, as well as the Vale of Peshawar just to the north of the Indus Valley (see also Chlenova 1984; Kuzmina 2007; P’yankova 1994). However, until recently, no Gandharan Grave culture sites or artifacts had been found in the region in between where BMAC and Vakhsh/Beshkek sites occur on the one hand (northern Afghanistan, southern Uzbekistan, and southern Tajikistan) and the Gandharan Grave culture (Lower Dir, Lower Swat, and the Vale of Peshawar) on the other. Here we may also encounter a lack of systematic archaeological surveys. However, anthropological analyses provide no support of a change of population (Hemphill 1998, 1999; Hemphill and Mallory 2004).

In 1968, Stacul 1969: 69 discovered a number of protohistoric cemetery sites near Chitral town, the capital of Chitral District, and he identified them as bearing close similarities to the Gandharan Grave culture sites reported further south. This conclusion was corroborated by Allchin 1970's study of three ceramic vessels recovered from the town of Ayun in southern Chitral. These too, were found to bear close affinities to vessels recovered from Gandharan Grave culture sites. In 1999, a joint Pakistani-British team carried out a survey in Chitral and recorded 15 cist graves identified as likely Gandharan Grave culture sites (Ali et al. 2002). This initial effort led to further survey and excavation in Chitral by a team of Pakistani archaeologists that resulted in the identification of additional large cemeteries and the excavation of a series of graves at the sites of Sangoor and Gankoreneotek, located near Chitral town (Ali et al. 2005b), and at Parwak, located near Mastuj (Ali et al. 2005a; Ali and Zahir 2005).

Radiocarbon dates obtained from three of the newly excavated Gandharan sites in Chitral District (Ali et al. 2008), which range from 1000 BCE to 1000 CE, are more recent than the age estimates for the lowland Gandharan sites (ca. 1700–500 BCE), confirming Stacul's (1970: 101) suspicion that the highland expressions of this technocomplex represent a subsequent development. Viewed as a whole, the evidence for contacts between the Gandharan Grave culture and any of the southern steppe cultures of the Late Bronze Age remain disputed and would demand more systematic archaeological coverage of the regions between the two groups as well as settlement evidence.

3.4. Concluding Remarks

This survey of the archaeological and biological record of southern Central Asia yields four important findings. First, contacts between the sedentary food-producing populations of the Namazga culture populations residing in Kopet Dagh piedmont and Geokyur oasis of southern Turkmenistan who likely established the outpost at Sarazm had little to no contact with populations residing in the southern steppe zone. Second, contacts between Bronze Age steppe populations and NMG V and BMAC populations appears to have been one in which the dynamic of cultural influence was stronger on the side of the well-established sedentary food-producing populations, and this resulted in the partial assimilation of these initial newcomers to the region both culturally and, to a lesser degree, biologically as well. Third, not all of those who emigrated from the north turned to farming but may have continued a semi-nomadic existence in the highlands, which were unsuitable for the kind of intensive farming practiced in the BMAC homelands or in the regions of Khorezm. Fourth, if there was any Central Asian influence on South Asian populations, that influence likely long predated any development of Iranian, let alone Indo-Aryan, languages, and most likely occurred during the late NMG IV to early NMG V period (ca. 2800–2300 BCE) and even earlier during the Eneolithic from Kelteminar culture groups (4000–3500 BCE).

4. Implications

The 4th and early 3rd millennia BCE mark a historical threshold linked to massive population growth and the rise of urban culture at the nexus of the African and Eurasian continents. Linked to this development were deep changes in technology as well as social and political organization. The period saw the establishment of the first complex long-distance trade networks, which in turn advanced a trafficking in commodities, ideas, and people. The following period has previously been portrayed as one of large-scale movement across Central Asia and further into South Asia, the so-called chariot horizon. Such movements have been linked to the mass movement of Indo-European-language speakers. In recent years, simple notions of mass migration and language spread have been contested and qualified on the basis of both material and written evidence, gradually being replaced by more complex models that combine migration, interaction, co-option, and conquest. The complexity of the situation in Bronze Age Anatolia, where questions of ethnicity and language can be addressed through both written and material sources, warns us that similarly complex conditions were probably in play in south-central Asia as well. Our hope to refine our understanding of actual population movement (a major feature of most historical models) through the use of genetics was the main stimulus behind this paper.

REFERENCES

Abdullaev, B. 1979. Mogilnik Dzharkutan. *Istorija Materialnoj Kultury Uzbekistana* 15: 1–22.

Alekshin, V. A. 1986. The Problem of the Origins of Archaeological Cultures of the Neolithic-Bronze Age in Central Asia (According to the Burial Rites). *Information Bulletin of the International Association for the Study of the Cultures of Central Asia* 11: 96–103.

Algaze, G. 1993. *The Uruk World System. The Dynamics of Expansion of Early Mesopotamian Civilization*. Chicago: University of Chicago Press.

Ali, I., C. Batt, R. Coningham, and R. Young. 2002. New exploration in the Chitral Valley, Pakistan: An extension of the Gandharan Grave culture. *Antiquity* 76: 647–53.

Ali, I., D. Hamilton, P. Newson, M. Qasim, R. Young, and M. Zahir. 2008. New radiocarbon dates from Chitral, NWFP, Pakistan, and their implications for the Gandharan Grave Culture of northern Pakistan. *Antiquity* 82. [http://www.antiquity.ac.uk/projgall/youngr/].

Ali, I., B. E. Hemphill, and M. Zahir. 2005a. Skeletal analysis of Gandharan graves at Parwak, Chitral (2003-04). *Frontier Archaeology* 3 183–226.

Ali, I., and M. Zahir. 2005. Excavation of Gandharan graves at Parwak, Chitral 2003-4. *Frontier Archaeology* 3: 135–82.

Ali, I., M. Zahir, and M. Qasim. 2005b. Archaeological survey of District Chitral, 2004. *Frontier Archaeology* 3: 91–106.

Allchin, F. R. 1970. A pottery group from Ayun, Chitral. *Bulletin of the School of Oriental and African Studies* 23(1): 1–4.

Amiet, P. 1986. *L'Age des Échanges Inter-Iraniens. 3500-1700 avant J.-C* (Notes et Documents des Musées de France No. 11). Paris: Musées de France.

Amiet, P. 1989. Antiquities of Bactria and Outer Iran in the Louvre Collection. In: G. Ligabue and S. Salvatori (eds.), *Bactra: An Ancient Oasis Civilisation from the Sands of Afghanistan*. Venice: Erizzo, 157–80.

Anthony, D. W. 2007. *The Horse, the Wheel and Language: How Bronze Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton: Princeton University Press.

Askarov, A. A. 1962. Pamjatniki Andronovskoi Kulturi v nizovyakh Zarafshana. *Istoriya Materialnoi Kultury Uzbekistana*, 3: 28–41.

Askarov, A. A. 1977. *Drevnezemledel'cheskaya Kul'tura Epokhi Bronzi Iuga Uzbekistana*. Tashkent: FAN.

Askarov, A. A. 1981. Southern Uzbekistan in the Second Millennium B.C. In: P. L. Kohl (ed.), *The Bronze Age Civilization of Central Asia: Recent Soviet Discoveries*. Armonk, NY: M.E. Sharpe, 256–72.

Askarov, A. A. and B. Abdullaev. 1983. *Dzharkutan*. Tashkent: FAN.

Avanesova, N. 1996. Pasteurs et Agriculturs del al Vallée du Zarafshan au début de l'Age du Bronze. In: B. Lyonnet (ed.), *Sarazm (Tadjikistan) Céramiques (Chalcolithique et Bronze Ancien)* (Mémoire de la Mission Archéologique Française en Asie Centrale 7). Paris: de Boccard.

Bakkels, C. C. 2003. The Contents of Ceramic Vessels in the Bactria-Margiana Archaeological Complex, Turkmenistan. *Electronic Journal of Vedic Studies*, 9(1c).

Barjamovic, G. 2011. *A Historical Geography of Anatolia in the Old Assyrian Colony Period* (Carsten Niebuhr Institute Publications 38). København: Museum Tusculanum.

Barjamovic, G. (in press). Interlocking Commercial Networks and the Infrastructure of Trade in Western Asia during the Bronze Age. In: K. Kristiansen, T. Lindkvist and J. Myrdal (eds.), *Trade and Civilization in the Pre-Modern World*. Cambridge: Cambridge University Press.

Batiuk, S. D. 2013 The fruits of migration: Understanding the 'longue dureé' and the socio-economic relations of the Early Transcaucasian culture. *Journal of Anthropological Archaeology* 32: 449–77.

Biscione, R. 1977. The Crisis of Central Asian Urbanisms in II Millennium B.C. and Villages as an Alternative System. In: J. Deshayes (ed.), *Le Plateau Iranien et l'Asia Centrale des Origines à la Conquête Islamique* (Colloques Internationaux 567). Paris: CNRS, 113–27.

Bobomulloev, S. 1997. Ein Bronzezeitliches Grab aus Zardca Chalifa bei Pentzekent (Zarafshan-Tal). *Archäologische Mittelungen aus Iran und Turan* 29: 121–34.

Boroffka, N., J. Cierny, J. Lutz, H. Parzinger, E. Pernicka, and G. Weisgerber. 2002. Bronze Age Tin from Central Asia: Preliminary Notes. In: K. Boyle, C. Renfrew and M. Levine (eds.), *Ancient Interactions: East and West in Eurasia* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 135–59.

Bostongukhar, S. 1998. *Verkov'ya Zarafshana vo II Tysyacheletii do n. er.* Dushanbe: Publisher.

Chlenova, N. L. 1984. Arkheologicheskie materialy k voprosu ob irantsakh doskifskoj epokhi I indo-irantsakh. *Sovetskaya Arkheologia* 1984(1): 88–103.

Derevyanko, A. P. and D. Dorj. 1992. Neolithic Tribes in Northern Parts of Central Asia. In: A. H. Dani and Masson (eds.), *History of Civilizations of Central Asia*. Paris: UNESCO, I: 163–83.

Dolukhanov, P. M. 1986. Farmers and Foragers in West-Central Asia. In: M. Zvelebil (ed.). *Hunters in Transition: Mesolithic Societies of Temperate Eurasia and Their Transition to Farming*. Cambridge: Cambridge University Press, 121–32.

Francfort, H.-P. 1979. *Les Fortifications en Asie Centrale de l'Age du Bronze à la l'Époque Kouchane* (Travaux de l'Unité de Recherches Archéoloques 10). Paris: Unité de Recherches Archéologiques.

Francfort, H.-P. 1981. The Late Periods of Shortughaï and the Problem of the Bishkent Culture (Middle and Late Bronze Age in Bactra). In: H. Härtel (ed.), *South Asian Archaeology 1979*. Berlin: Dietrich Reimer Verlag, 191–202.

Francfort, H.-P. 1984. The Early Periods of Shortughai (Harappan) and the Western Bactrian Culture of Dashly. In: B. Allchin (ed.), *South Asian Archaeology 1981*. Cambridge: Cambridge University Press, 170–75.

Francfort, H.-P. 1992. Dungeons and Dragons: Reflections on the System of Representations in Protohistoric Bactria and Margiana. In: G. L. Possehl (ed.), *South Asian Archaeology Studies*. Delhi: Oxford and IBH Publishing, 179-208.

Francfort, H.-P. 1994. The Central Asian Dimension of the Symbolic System in Bactria and Margiana. *Antiquity* 68: 406–18.

Francfort, H.-P. 2001. The cultures with painted ceramics of south Central Asia and their relations with the northeastern steppe zone (late 2nd-early 1st millennium BC). In: R. Eichmann and H. Parzinger (eds.), *Migration und Kulturtransfer. Der Wandel vorder-und zentralasiatischer Kulturen im Umbruch vom 2. zum 1. vorchristlichen Jahrtausend*. Bonn: Habelt.

Francfort, H.-P. 2005. La civilisation de l’Oxus et les Indo-Iraniens et Indo-Aryens en Asie centrale. In: G. Fussman, J. Kellens, H.-P. Francfort, and X. Tremblay (eds.), *Āryas, Aryens et iraniens en Asie central*. Paris: Diffusion De Boccard.

Franke-Vogt, U. 2001 The southern Indus Valley during the later 2nd and 1st millennia BC. The Dark Age. In: R. Eichmann and H. Parzinger (eds.), *Migration und Kulturtransfer. Der Wandel vorder- und zentralasiatischer Kulturen im Umbruch vom 2. zum 1. vorchristlichen Jahrtausend*. Bonn: Habelt Verlag.

Götzelt, T. 1995. Complex Simplicities: Of Periods and Functions in Bronze Age Southern Turkmenistan. *Altorientalische Forschungunen* 22(1): 56–69.

Gulyamov, G., U. I. Islamov, and A. A. Askarov. 1966. *Pervobitnaya Kul’tura i Vozniknoveniye Oroshayemogo zemledeliya v nizov’yakh Zarafshana*. Tashkent: FAN.

Gupta, S. P. 1979. *Archaeology of Soviet Central Asia, and the Indian Borderlands*. Delhi: B. R. Publishing Corporation.

Hanks, B. K., A. V. Epimakhov, and A. C. Renfrew. 2007. Towards a Refined Chronology for the Bronze Age of the Southern Urals, Russia. *Antiquity* 81: 353–67.

Hansen, S. 2010. Communication and exchange between the Northern Caucasus and Central Europe in the fourth millennium BC. In: S. Hansen, A. Hauptmann, I. Motzenbäcker, and E. Pernicka (eds.), *Von Maykop bis Trialeti. Gewinnung und Verbreitung von Metallen und Obsidian in Kaukasien im 4.-2. Jt. V. Chr*. Bonn: Habelt Verlag, 297–316

Hansen, S. 2014 The 4th Millennium: A Watershed in European Prehistory. In: B. Horejs and M. Mehofer (eds.), *Western Anatolia before Troy. Proto-Urbanisation in the 4th Millennium BC?* Vienna: Verlag der Österreichischen Akademie der Wissenschaften, 243–59.

Harmatta, J. 1992. The Emergence of the Indo-Iranians: The Indo-Iranian Languages. In: A. H. Dani and V. M. Masson (eds.), *History of Civilizations of Central Asia*. Paris: UNESCO, I: 346–70.

Hemphill, B. E. 1998. Biological affinities and adaptations of Bronze Age Bactrians: III. An initial craniometric assessment. *American Journal of Physical Anthropology* 106: 329–48.

Hemphill, B. E. 1999a. Foreign Elites from the Oxus Civilization?: A Craniometric Study of Anomalous Burials From Bronze Age Tepe Hissar. *American Journal of Physical Anthropology* 110: 421–34.

Hemphill, B. E. 1999b. Adaptations and Affinities of Bronze Age Bactrians: IV. A Craniometric Examination of the Origins of Oxus Civilization Populations. *American Journal of Physical Anthropology* 108: 173–92.

Hemphill, B. E. 2013. Grades, Gradients, and Geography: A Dental Morphometric Approach to the Population history of South Asia. In: G. R. Scott and J. D. Irish (eds.), *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation*. Cambridge: Cambridge University Press, 341–87.

Hemphill, B. E. and J. P. Mallory. 2004. Horse-Mounted Invaders from the Russo-Kazakh Steppe or Agricultural Colonists from Western Central Asia? A Craniometric Investigation of the Bronze Age Settlement of Xinjiang. *American Journal of Physical Anthropology* 124(3): 199–222.

Hiebert, F. T. 1994. *Origins of the Bronze Age Civilization in Central Asia* (American School of Prehistoric Research Bulletin No. 42). Cambridge, MA: Peabody Museum of Archaeology and Ethnology.

Hiebert, F. T. 2002a. The Kopet Dag Sequence of Early Villages in Central Asia. *Paléorient* 28(2): 25–42.

Hiebert, F. T. 2002b. Bronze Age Interaction between the Eurasian Steppe and Central Asia. In: K. Boyle, A. C. Renfrew, and M. Levine (eds.). *Ancient Interactions: East and West in Eurasia* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 237–48.

Hiebert, F. T. and D. Killick. 1993. Metallurgy of Bronze Age Gonur Depe. *Information Bulletin of the International Association for Study of the Cultures of Central Asia* 19: 186–204.

Isakov, A. I. and T. M. Potemkina. 1989. Mogil’nik plemon Epokhi Bronzi v Tadzhikistane. *Sovetskaya Arkheologiya* 1: 145–69.

Ivanova, M. 2012 Kaukasus und Orient: Die Entstehung des „Maykop-Phänomens“ im 4. Jahrtausend v. Chr. *Praehistorische Zeitschrift* 87(1): 1–28.

Jarrige, J.-F. 1989. Funerary Complexes in Baluchistan at the End of the Third Millennium in the Light of Recent Discoveries at Mehrgarh and Quetta. In: K. Frifelt and P. Sørensen (eds.), *South Asian Archaeology 1985* (Scandinavian Institute of Asian Studies Occasional Papers 4). London: Curzon Press, 150–60.

Jarrige, J.-F. 1994. The Final Phase of the Indus Occupation at Nausharo and its Connection with the following Cultural Complex of Mehrgarh VIII. In: A. Parpola and P. Koskikallio (eds.), *South Asian Archaeology 1993* (Annales Academiae Scientarum Fennicae 271, Series B, Volume I). Helsinki: Suomalaisen Tiedeakatemian Toimituksia, 295–314.

Jarrige, J.-F. and M. U. Hassan. 1989. Funerary Complexes in Baluchistan at the End of the Third Millennium in the Light of Recent Discoveries at Mehrgarh and Quetta. In: Frifelt, K. and P. Sørensen, P. (eds.), *South Asian Archaeology 1985* (Scandinavian Institute of Asian Studies Occasional Papers No. 4). London: Curzon Press. Scandinavian Institute of Asian Studies, 150–66.

Khlopin, I. N. 1989. Origins of the Bronze Age Cultures in South Central Asia. *Information Bulletin of the International Association for the Study of the Cultures of Central Asia* 15: 74–84.

Khlopin, I. N. 1994. Antiquities of southwestern Turkmenistan. In: A. Parpola and P. Koskikallio (eds.). *South Asian Archaeology 1993* (Annales Academiae Scientarum Fennicae 271, Series B, Volume I). Helsinki: Suomalaisen Tiedeakatemian Toimituksia, 363–67.

Khlopina, L. I. 1972. Southern Turkmenia in the Later Bronze Age. *East and West* 22: 199–214.

Kohl, P. L. 1984. *Central Asia. Paleolithic Beginnings to the Iron Age* (Synthése 14). Paris: Éditions Recherche sur les Civilisations.

Kohl, P. L. 1992. Central Asia (Western Turkestan): Neolithic to the Early Iron Age. In: R. W. Erich (ed.), *Chronologies in Old World Archaeology*. Chicago: University of Chicago Press, I: 179–96.

Kohl, P. L. 2002. Comment on Lamberg-Karlovsky's "Archaeology and Language." *Current Anthropology* 43(1): 77–78.

Kohl, P. L. 2007. *The Making of Bronze Age Eurasia*. Cambridge University Press: Cambridge.

Kozintsev, A. G. 2000. On the Biological Affinities and Origin of the Northern Pontic Scythians. *Archaeology, Ethnology, and Anthropology of Eurasia* 3(3): 145–52.

Kozintsev, A. G. 2008. The “Mediterraneans” of Southern Siberia and Kazakhstan, Indo-European Migrations, and the Origins of the Scythians: A Multvariate Craniometric Analysis. *Archaeology, Ethnology and Anthropology of Eurasia* 36(4): 140–44.

Kozintsev, A. G. 2010. Craniometric Evidence of the Early Caucasoid Migrations to Siberia and Eastern Central Asia, with Reference to the Indo-European Problem. *Archaeology, Ethnology and Anthropology of Eurasia* 37(4): 125–36.

Kristiansen, K., and Larsson, T. B., 2005. *The Rise of Bronze Age Society*. Travels, Transmissions and Transformations. Cambridge University Press.

Kroonen, G., G. Barjamovic, and M. Peyrot. 2018. Linguistic supplement to Damgaard et al. 2018: Early Indo-European Languages, Anatolian, Tocharian and Indo-Iranian. *BioR* 14 DOI XXX.

Kuzmina, E. E. 2007. *The Origin of the Indo-Iranians*, ed. J. P. Mallory.
Leiden/Boston: Brill.

Kuzmina, E. E. 1986. *Drevneishie skotovodi ot Urala do Tyan-Shanya*. Frunze: Ilium.

Kuzmina, E. E. 2001. The first wave of Indo-Iranians to the South. *Journal of Indo-European Studies* 29(1–2): 1–40.

Kuzmina, E. E. 2003. Origins of Pastoralism in the Eurasian Steppes. In: M. Levine, A. C. Renfrew, and K. Boyle (eds.), *Prehistoric Steppe Adaptation and the Horse* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 203–32.

Kuzmina, E. E. 2007. *The Origin of the Indo-Iranians*. Leiden: Brill.

Lamberg-Karlovsky, C. C. 2002. Archaeology and Language. *Current Anthropology* 43(1): 63–88.

Larsen, M. T. 2015 *Ancient Kanesh. A Commercial Colony in Bronze Age Anatolia*. Cambridge: Cambridge University Press.

Larsen, M. T. and A. W. Lassen. 2014. Cultural Exchange at Kültepe. In: M. Kozuh (ed.), *Extraction and* Control. Studies in Honor of Matthew W. Stolper (Studies in Ancient Oriental Civilization 68). Chicago: University of Chicago, 170–88.

Lassen, A. W. 2010. The trade in wool in Old Assyrian Anatolia. *Jaarbericht ‘Ex Oriente Lux’* 42: 159–79.

Laursen, S. and P. Steinkeller. 2017. *Babylonia, the Gulf Region and the Indus: Archaeological and Textual Evidence for Contact in the Third and Early Second Millenia BC*. Winona Lake, IN: Eisenbrauns.

Lightfoot, E., G. Motuzaite-Matuzeviciute, T. C. O’Connell, I. A. Kukushkin, V. Loman, V. Varfolomeev, X. Liu, and M. K. Jones. 2015. How ‘Pastoral’ is Pastoralism? Dietary Diversity in Bronze Age Communities in the Central Kazakhstan Steppes. *Archaeometry* 57 Suppl. 1: 232–49.

Litvinsky. B. A. 1962. Rudnaya baza, metallurgiya i liteinoe delo. In: B. A. Litvinsky, A. P. Okladnikov, and V. A. Ranov (eds.). *Drevnosti Kairak-Kumov (Drevneishaya istoriya Severnovo Tadzikistana)*. Dushanbe: Akademija Nauk Tadzhikskoi SSR, 170–95.

Lyonnet, B. 1994. Central Asia, the Indo-Aryans and the Iranians: Some reassessments from Recent Archaeological Data. In: A. Parpola and P. Koskikallio (eds.). *South Asian Archaeology 1993* (Annales Academiae Scientarum Fennicae 271, Series B, Volume I). Helsinki: Suomalaisen Tiedeakatemian Toimituksia, 425–34.

Lyonnet, B., T. Akhundow, K. Almamedow, L. Bouquet, B. Jellinov, F. Huseynov, S. Loute, Z. Makharadze, and S. ReynardS. 2008 Late Chalcolithic Kurgans in Transcaucasia. The cemetery of Soyuq Bulaq, (Azerbaijan). *Archäologische Mitteilungen aus Iran und Turan* 40: 27–44.

Masson, V.M. 1992a. The Bronze Age in Khorasan and Transoxiana. In: A. H. Dani and V. M. Masson (eds.), *History of Civilizations of Central Asia*. Paris: UNESCO, I: 225–45.

Masson, V. M. 1992b. The Decline of the Bronze Age Civilization and Movements of the Tribes. In: A. H. Dani and Masson, V.M. (Eds.), *History of Civilizations of Central Asia*. Paris: UNESCO, I: 326–45.

Masson, V. M. and T. P. Kiiatkina. 1981. Man at the Dawn of Civilization. In: P. L. Kohl (ed.), *The Bronze Age Civilization of Central Asia: Recent Soviet Discoveries*. Armonk, NY: M.E. Sharpe, 107–34.

Michel, C. 2004. The Perdum-Mule, a Mount for Distinguished Persons in Mesopotamia during the First Half of the Second Millennium BC. In: B. S. Frizell (ed.), *PECUS. Man and Animal in Antiquity. Proceedings of the Conference at the Swedish Institute in Rome, September 9-12, 2002*. (Projects and Seminars 1). Rome: Swedish Institute, 190–99.

Narimanishvili, G. and N. Shanshashvili. 2010. Neue Forschungen zur Trialeti-Kultur. In: S. Hansen, A. Hauptmann, I. Motzenbäcker and E. Pernicka (eds.), *Von Maykop bis Trialeti. Gewinning und Verbreitung von Metallen und Obsidian in Kaukasien im 4.-2. Jt. V. Chr*.. Bonn, Habelt Verlag, 205–22.

Negmatov, N. 1982. The Development of Urbanized Cultures in Tajikistan and the Problem of Culture Contacts. *Information Bulletin of the International Association for the Study of the Cultures of Central Asia* 2: 61–63.

P’yankova, L. 1989. Pottery Complexes of Bronze Age Margiana (Gonur and Togolok-21). *Information Bulletin of the International Association for the Study of the Cultures of Central Asia* 16: 27–54.

P’yankova, L. 1993. Pottery of Margiana and Bactria in the Bronze Age. *Information Bulletin of the International Association for the Study of the Cultures of Central Asia* 19: 109–27.

P’yankova, L. 1994. Central Asia in the Bronze Age: Sedentary and Nomadic Cultures. *Antiquity* 68: 355–72.

Palumbi, G. 2011. The Arslantepe royal tomb and the “Manipulation” of the kurgan ideology in eastern Anatolia at the beginning of the third millennium. In: E. Borgna and S. M. Celka (), *Ancestral Landscapes* (TMO 61). Lyon: Maison de l’Orient and de la Méditerannée, , 47–59.

Parpola, A. 2015. *The Roots of Hinduism. The Early Aryans and the Indus Civilization*. Oxford: Oxford University Press.

Parpola, A. 1988. The Coming of the Indo-Aryans to Iran and India and the Cultural and Ethnic Identity of the Dasas. *Studia Orientalia* 64: 195–302.

Parpola, A. 1993. Margiana and the Aryan Problem. *Information Bulletin of the International Association for the Study of the Cultures of Central Asia* 19: 41–62.

Parpola, A. 1995. Formation of the Aryan Branch of Indo-European. In: R. Blench and M. Spriggs (eds.), *Language and Archaeology, Volume 3: Combining Archaeological and Linguistic Aspects of the Past*. London: Routledge, 1–27.

Parzinger, H. 2003. Zinn in der Bronzezeit Eurasiens. In: H. Parzinger and N. Boroffka (eds.), *Das Zinn der Bronzezeit in Mittelasien I. Die Siedlungsarchäologischen Forschungen im Umfeld der Zinnlagerstellen* (Archäologie in Iran und Turan 5). Mainz am Rhein: Verlag Philipp von Zabern, 287–96.

Pratt, M. L. 1992. *Imperial Eyes: Travel Writing and Transculturation*. London: Routledge.

Puturidze, M. 2016. Some evidence for south Caucasia Aegean world relationship. In: M. Kvachadze and N. Shanshashvili (eds.), *Aegean World and South Caucasus. Cultural Relations in the Bronze Age*. Tbilisi: Publisher, 139–62.

Rahmsdorf, L. 2010. Indications of Aegean-Caucasian relations during the third millennium BC. In: S. Hansen, A. Hauptmann, I. Motzenbäcker, and E. Pernicka (eds.), *Von Maykop bis Trialeti. Gewinnung und Verbreitung von Metallen und Obsidian in Kaukasien im 4.-2. Jt. V. Chr.*. Bonn: Habelt Verlag, 263–95.

Ratnagar, S. 2006. *Trading Encounters: From the Euphrates to the Indus in the Bronze Age*, 2nd ed. Oxford: Oxford University Press.

Rubinson, K. 2003. Silver Vessels and Cylinder Sealings: Precious Reflections of Economic Exchange In the Early Second Millennium BC. In: A. T. Smith and Rubinson, K.S. (eds.), *Archaeology in the Borderlands: Investigations in Caucasia and Beyond* Los Angeles: Cotsen Institute of Archaeology, 128–43.

Sagona, A. 2011. Anatolia and the Transcaucasus: Themes and Variations ca. 6400–1500 B.C.E. In: S. R. Steadman and G. McMahon (eds.), *The Oxford Handbook of Ancient Anatolia*. Oxford/New York: Oxford University Press, 683–703.

Salvatori, S. 1994. A Late Third Millennium Graveyard at Gonur Tepe 1 (Murghab Delta, Turkmenistan). In: A. Parpola and P. Koskikallio (eds.), *South Asian Archaeology 1993* (Annales Academiae Scientarum Fennicae 271, Series B, Volume I). Helsinki: Suomalaisen Tiedeakatemian Toimituksia, 657–66.

Salvatori, S. 1995. Protohistoric Margiana: On a Recent Contribution. (Review of: “IASCCA (International Association for the Study of the Cultures of Central Asia) Information Bulletin” 19, Moscow 1993). *Revista di Archeologia* 19: 38–55.

Salvatori, S. 2008. Cultural Variability in the Bronze Age Oxus Civilisation and its relations with the Surrounding Regions of Central Asia and Iran. In: S. Salvatori, M. Tosi, and B. Cerasetti (eds.), *The Bronze Age and Early Iron Age in the Margiana Lowlands: Facts and Methodological Proposals for a Redefinition of the Research Strategies*. Oxford: Archaeopress, 75–98.

Salvatori, S. 2016. Bactria-Margiana Archaeological Complex: how terminology hides historical processes. In: N. A. Dubova (ed.), *Transactions of Margiana Archaeological Expeditions, Volume 6. To the Memory of Professor Victor Sarianidi*. Moscow: Altai State University, 449–60.

Salvatori, S., M. Vidale, G. Guida, and G. Gigante. 2003. A Glimpse on Copper and Lead Metalworking at Atyn-depe (Tukrmenistan) in the 3rd Millennium BC. *Ancient Civilizations from Scythia to Siberia* 8(1–2): 69–106.

Santoni, M. 1984. Sibri and the South Cemetery at Mehrgarh. Third Millennium Connections between the Northern Kachi Plain (Pakistan) and Central Asia. In: B. Allchin (ed.), *South Asian Archaeology 1981*. Cambridge: Cambridge University Press, 52–60.

Sarianidi, V. I. 1979. K voprosu o Kult’ure Zamanbaba. *Ethnografiya i Arkheologiya Srednei Azii* 23–28.

Sarianidi, V. I. 1981. Margiana in the Bronze Age. In: P. L. Kohl (ed.), *The Bronze Age Civilization of Central Asia. Recent Soviet Discoveries*. Armonk, NY: M.E. Sharpe, 165–93.

Sarianidi, V. I. 1990. *Drevnosti strain Margush*. Ashkkabad: Akademia Nauk Turkmenskoi.

Sarianidi, V. I. 1993. Margiana in the Ancient Orient. *International Association for the Study of the Cultures of Central Asia* 19: 5–28.

Sarianidi, V. I. 1998. *Margiana and Protozoroastrianism*. Athens: Kapon Editions.

Sarianidi, V. I. 1999. Near Eastern Aryans in Central Asia. *Journal of Indo-European Studies*, 27(3–4): 295–326.

Sarianidi, V. I. 2008. Korolevsky Nekropol’ v Severnom Gonur. *Vestnik Drevnej Istorii* 2006(2): 155–92.

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

Spengler, R. N., B. Cerasetti, M. Tengberg, M. Cattani, and L. M. Rouse. 2014. Agriculturalists and pastoralists: Bronze Age economy. *Vegetation History and Archaeobotany* 23: 805–20. DOI: 10.1007/s00334-014-0448-0.

Stacul, G. 1969. Excavation near Ghatigai (1968) and chronological sequence of proto-historical cultures in the Swat Valley. *East and West* 19: 44–91.

Stacul, G. 1970. The gray pottery in the Swat Valley and the Indo-Iranian connections (ca. 1500-300 B.C.). *East and West* 20: 92–102.

Terekhova, N. N. 1981. The History of Metalworking Production among the Ancient Agriculturalists of Southern Turkmenia. In: P. L. Kohl (ed.). *The Bronze Age Civilizations of Central Asia: Recent Soviet Discoveries*. Armonk, NY: M.E. Sharpe, 313–24.

Tolstov, S. P. 1962. *Po drevnim del’tam Oksa i Jaksarta*. Moscow: Nauka.

Tolstov, S. P., T. A. Zdanko, and M. A. Itina. 1963. Raboti Chorezmskoi Archeologo-Etnografisheskoi Ekspedicii AN SSR v 1958-1961gg. *Materialy Chorezmskoi Ekspedicii* 6: 3–90.

Vinogradova, N. 1993. Interrelation between Farming and ‘Steppe’ Tribes in the Bronze Age South Tadjikistan. In: A. J. Gail and G. J. R. Mevissen (eds.), *South Asian Archaeology 1991*. Stuttgart: Franz Steiner Verlag, 289–301.

Vinogradova, N. and E. E. Kuzmina. 1986. Kontakti Stepnykh i Zemledel’cheskikh plemen Srednei Azii v Ephoki Bronzi. In: B. A. Litvinsky (ed.), *Vostochnyi Turkestan i Srednyaya Aziya*. Moscow: Nauka, 126–51.

Wilkinson, T. C. (in press). Cloth and currency: on the ritual-economics of Eurasian textile circulation and the ‘origins’ of trade, fifth to second millennia BC. In: K. Kristiansen, T. Lindkvist, and J. Myrdal (eds.), *Trade and Civilisation*. Cambridge: Cambridge University Press.

Wright, R. P. 2010. *The Ancient Indus. Urbanism, Economy, and Society* (Case Studies in Early Societies). Cambridge: Cambridge University Press.

Zahir, M. 2016. The “Gandhara Grave Culture.” In: G. R. Schlug and S. R. Walimbe (eds.), *A Companion to South Asia in the Past*. New York: Wiley & Sons, 274–93.

Zych, L. 2006. Perspectives on Chalcolithic Turkmenistan: The Global and the Local in the Kopet Dag. In: D. L. Peterson, L. M. Popova, and A. T. Smith (eds.), *Beyond the Steppe and the Sown. Proceedings of the 2002 University of Chicago* (Colloquia Pontica 13). Leiden/Boston: Brill, 121–21.

Archaeological supplement B to Damgaard et al. 2018: discussion of the archaeology of Central Asian and East Asian Neolithic to Bronze Age hunter-gatherers and early pastoralists, including consideration of horse domestication.

AUTHORS

Alan K. Outram[1], Alexey Polyakov[2], Andrei Gromov[3], Vyacheslav Moiseyev[3], Andrzej W. Weber[4], Vladimir I. Bazaliiskii[5], O. I. Goriunova[5,6]

AFFILIATIONS

[1]Department of Archaeology, University of Exeter, Exeter, EX4 4QE, UK.

[2]Institute for the History of the Material Culture, Russian Academy of Sciences.

[3]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, Russia.

[4]Department of Anthropology, University of Alberta, Edmonton, Alberta, T6G 2H4, Canada.

[5]Department of History, Irkutsk State University, Karl Marx Street 1, Irkutsk 664003, Russia.

[6]Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, Academician Lavrent'iev Ave. 17, Novosibirsk, 630090, Russia.

ABSTRACT

The archaeological evidence relating to selected key cultures from Central and East Asia from the Neolithic to the Bronze Age is summarized. These cultures include the Eneolithic (Copper Age) Botai culture of northern Kazakhstan, the Bronze Age Okunevo culture from the Minusinsk Basin in Russia and Neolithic to Bronze Age cultures of the Baikal Region in East Siberia. Special consideration is given to the debate surrounding horse domestication within the Botai Culture, and the key lines of evidence are summarized.

1. Horse Domestication and the Botai Culture (Alan K. Outram)

1.1 Horse Domestication in the Central Asian Steppe:

The domestication of the horse is widely recognized as being of immense importance to the development of human societies, revolutionizing transport, trade, and modes of warfare (Anthony 2007; Olsen 2006; Outram et al. 2009). Recently, however, a number of large-scale analyses of human ancient DNA suggest that the development of mobile pastoral societies in the Eurasian steppe, particularly the Yamnaya culture of the Pontic Steppe, was responsible for a major period of human migration into Europe around 5,000 years ago that may well be related to the arrival of Indo-European languages and culture in Europe (Allentoft et al. 2015; Haak et al. 2015). The development of these societies has been linked to horse riding, mixed herding, the use of wheeled transportation, and bronze metallurgy (Anthony and Ringe 2015). As such, understanding the earliest development of horse husbandry and pastoral economic systems in the steppes of Eurasia must be regarded as one of the big questions in prehistoric archaeology. Following the arrival of agriculture, this development arguably marks the beginnings of the next major phase of Anthropocene impacts on the environment, with vastly increased mobility representing the incipient phases of globalization, since Central Asia is a continental crossroads containing crucial East-West trade routes, potentially highly significant in initial "Trans-Eurasian Exchange" (Jones et al. 2011; Sherratt 2006).

With the exception of the dog, the reindeer, and South American camelids, it seems that animal domestications were generally undertaken by farmers (Outram 2014). The domestication of cattle, sheep, goats, and pigs in the Near East appears to have happened after a significant period during which the economy relied upon cereal agriculture alongside the hunting of wild gazelle, while in most other centers of domestication animals were, at best, domesticated at the same time as plants were (Outram 2014). Dog domestication is the earliest animal domestication, being clearly undertaken by people of the Palaeolithic (Sablin and Khlopachev 2002; Savolainen et al. 2002; Wayne et al. 2006). It is anomalous because, while dogs could be eaten at times, the relationship was much more likely to be related to mutual benefit with regard to hunting (Outram 2014). This is the classic example of

Zeder 2012’s “commensal pathway” to domestication. The early phases of reindeer domestication are poorly understood, but this must clearly have followed Zeder 2012’s “prey pathway” to domestication. Unusually, it is an example of a hunter-gatherer population changing its long-standing hunting relationship to one of herding, rather than domestication, by an expanding farming population that was putting pressure on wild animal resources. Zeder (2012, p176) has suggested that horses might represent an example of the “direct pathway” to domestication, where domestication is a “intention-driven, directed process.” It is essential to understand the origins of the Botai people in order to establish the likely domestication route. Directed domestication implies a prior understanding of the concept, so it would be more likely to be true if the Botai had origins among people with familiarity of herding and stock raising. Yet, if it were a local adaption by hunter-gatherers familiar with horse hunting for millennia, then this would be a unique example of a very late hunter-gatherer “prey pathway” domestication—but one that had the potential of massive effects upon human societies once horses were harnessed as well as eaten. A further key question must relate to the nature of the relationship between Botai, their domestic horses, and peoples such as the Yamnaya.

There are two major ecological zones within northern and central Kazakhstan. In the north there is the “forest steppe,” made up of a patchwork of grassland with stands of birch and pine trees, while in the central region there is a relatively treeless, semi-arid steppe. The area was a steppe in prehistory also, though there was variation over time in relation to tree cover, with pine generally increasing in extent from the 4th millennium BCE through to the Iron Age (Kremenetski et al. 1997). Significant cereal agriculture appears not to have been practiced in the region until the Soviet period. The Neolithic of northern and central Kazakhstan (so-called because it possessed ceramics) appears to have had an economy based upon hunting, gathering, and fishing, and its stone tool tradition consisted mainly of blade technology. With a few exceptions, settlements are rather ephemeral, and many comprise little more than scatters of material with no solid evidence for farming activities. Around 3500 BCE, northern Kazakhstan sees a new phenomenon with the Botai culture manifesting major changes in economic focus, settlement structure, and material culture (Zaibert 2009). Pottery use becomes more widespread, and lithic technologies change to bifaces and ground stone tools. The Botai Culture develops sizeable settlements that can have more than 100 semi-subterranean pit houses. Whether these were seasonally used or sedentary sites is not currently known. The most significant change, however, is a sudden and extreme focus on the exploitation of horses. Horse bones represent the vast majority of faunal assemblages at all Botai sites, and at Botai itself they reach the level of 99% of the faunal assemblage (Olsen 2006). The steppes of Central Asia had a substantial population of wild horses that were also available to earlier prehistoric groups in the region as a prey animal. With Botai, however, one sees a sudden focus on that animal, in conjunction with the arrival of substantial villages and significant changes in material culture. Since Botai was discovered in the early 1980s, there has been considerable discussion over whether the horses were hunted or herded and whether they were biologically domestic or still wild. Some have argued that there was no clear size change in the animals (Benecke and von den Driesch 2003) and that there was not a clearly managed herd structure for meat production (Levine 2004). However, size change need not be an immediate consequence of all domestication events, and herd structures would not be optimized for meat if horses were also being exploited for secondary products such as milk, riding, or traction (Anthony and Brown 2011; Outram 2014). Others have argued that the nature of the settlements and the low frequency of hunting material culture suggested control of the horse population and that multiple uses of horses for food and riding resulted in the broad herd structures seen (Olsen 2006). There is also established evidence for riding in the form of pathological bit-wear traces on the lower second premolars (Brown and

Anthony 1998) in a form now known as type 1 bit wear manifested in as a beveled facet on the tooth (Anthony and Brown 2011).

Following further recent investigations (Outram et al. 2009) it is now clear that at least some of the Botai horses were herded and domestic. This new study confirmed evidence of bit-wear and harnessing pathologies using different but complementary techniques (Outram et al. 2009), known as type 2 (parallel band of wear down the front of the 2nd mandibular premolar) and type 3 bit wear (pathology of diastema) (Anthony and Brown 2011). Furthermore, Botai pottery contained two types of equine lipid residues identified as adipose fat and mare’s milk fat (Outram et al. 2009), providing a clear indication of animal husbandry and secondary products use. Genetic research had also suggested that the date and general region of Botai fit with evidence for an increase in the frequency of coat colors in horses that are normally very rare in the wild and thus likely the result of domestic management (Ludwig et al. 2009). Indeed recent study of ancient genomes from the Botai horses themselves has also identified the significant presence of the leopard-spotting complex. This coat color is associated with human husbandry and selection in early domestic horses, and such control could have been exerted at Botai through the use of corrals that have now been archaeologically evidenced at more than one Botai culture site. Importantly, however, this study also concludes that Botai horses are not the principal source of modern domestic horse stock (Gaunitz et al. 2018). While earlier events of horse domestication remain possible and at least one other center of domestication is likely, Botai currently still represents the earliest unambiguous evidence for the herding and riding of domestic horses (Anthony and Brown 2011).

As such, it seems likely that early pastoralism in the region may have started with the horse but without arable agriculture, and it encompassed secondary as well as primary products. The Botai culture ends at the start of the 3rd millennium BCE. The following Early Bronze Age (c. 3,000-2,200 BCE) in that region shows the arrival of mixed pastoralism, with the addition of domestic cattle, sheep, and goats (Frachetti 2008). At this same time, the Yamnaya culture of the Pontic-Caspian steppe sees rapid territorial expansion up the Danube, making use of cattle, horses, and wheeled vehicles (Anthony 2007). The timing of this development, following evidence of horse domestication in the adjacent Central Asia Steppe, is unlikely to be coincidental, but the relationship between Botai and Yamnaya is in need of further investigation.

1.2 Botai Culture Origins:

A very significant question about the Botai culture is whether it was a local development from preceding Neolithic hunter-gatherer cultures, the result of inward migration, or a combination of local culture with outside influences. The immediately preceding Neolithic cultures in northern Kazakhstan were the Atbasar and Makhandzhar cultures (Kislenko and Tatarintseva 1999). Atbasar centers around the river Ishim, while Makhandzhar around the river Tobol. While possessing ceramics, hence their Neolithic label, their economy was based upon hunting and gathering in the forest steppe, and probably also fishing. Neolithic lithic technology focused strongly on blade production whereas the later Eneolithic cultures such as Botai made considerable use of bifacially-flaked stone technology (Kislenko and Tatarintseva 1999). While the ceramic tradition of the Botai is not radically different from the preceding Neolithic, the change in lithic technology is significant.

Kislenko and Tatarintseva (1999) suggest that the Atbasar and Makhandzhar were involved in the development of the Botai culture but under influences coming from the eastern Caspian and southern Urals. This explanation allows for adaptation of local peoples influenced by external cultural ideas. Such an origin from local, hunter-gatherer Neolithic peoples is also favored by Botai’s original and long-term investigator, Victor Zaibert (Zaibert 2009). On the other hand, scholars such as Anthony (Anthony and Brown 2011) suggest significant influence from migrating peoples from the

Volga-Ural steppes in the genesis of the Botai culture in northern Kazakhstan and, later, the Afanasievo culture in the Altai. The former solution would suggest a local, hunter-gatherer genetic origin for the Botai, while the latter suggests genetic influx from more westerly pastoralist groups, perhaps resulting in admixture. The former lends itself to an original domestication event based upon the “prey pathway,” while the latter suggests either “directed” domestication of a local species by people familiar with herding or introduction of domestic horses from outside.

1.3 The Botai Site:

Excavations have been conducted at the Eneolithic settlement of Botai under the direction of Victor Zaibert since 1980. The site dates to the mid- to late 4th millennium BCE (Levine and Kislenko 2002; Outram et al. 2009) and is the type site for a wider culture that includes a number of similar settlements, the most important of which are Krasnyi Yar and Vasilkovka (Olsen et al. 2006). A key feature of all these sites is the extreme dominance of horses in their faunal assemblages, almost to the exclusion of other species (Olsen et al. 2006). Ever since these sites were discovered, therefore, Botai has been the focus of many discussions about early horse domestication, herding, and riding. Botai culture sites consist of a very significant number of houses arranged in long rows, as seen at Krasnyi Yar and Vasilkovka (Olsen et al. 2006), or both rows and circular clusters, as seen at Botai itself (Gaunitz et al. 2018). The houses are sub-circular pit houses dug about 1 m below the ground surface and between about 5-8 m across. Their floors are compressed, clay-rich soil, and there are usually fairly central hearth pits, plus occasional eccentric storage pits, but no clear evidence for the precise nature of roofing or roof support. The houses are generally ringed by pits that are rich in bone deposits that are heavily dominated by horses (Olsen et al. 2006; Zaibert 2009; Zaibert et al. 2007), but usually there is also a dog burial or cranium in at least one associated pit (Olsen 2000).

Human burials are very rare in the Botai culture (Olsen 2006), and only a very small number of features containing human remains have been found, and all of these are at the site of Botai itself. The most significant of these features was a large pit that contained the remains of 4 individuals (2 adult men, an adult woman, and a 10–11-year-old child) along with the partial remains of at least 14 horses, principally crania, that formed an arc around the edges of the pit (Olsen 2006; Zaibert 2009). In 2005, a partially disarticulated inhumation was also discovered that lacked significant accompanying deposits (Zaibert et al. 2007). In addition to these inhumations two disarticulated human crania have also been found—one had a clay mask applied to it before it was buried in a pit outside a house, and the other had been made into a bowl (Olsen 2006). Most recently, in 2016, a further almost complete individual was found in a shallow grave next to a house in an unusual posture without any identifiable funerary rite or grave goods. It is clear, from this evidence, that we currently lack a sound understanding of Botai culture funerary practices, and these few inhumations may not be “normative” in nature. Archaeological exploration has been concentrated on the settlements themselves, and currently there are not obviously recognizable monuments or surface finds that might indicate the presence of accompanying cemeteries. What is clear is that horses were an important part of Botai culture ritual deposits, along with dogs, and that skulls, whether human or animal, held particular significance.

2. Okunevo (Alexey Polyakov, Andrei Gromov, Vyacheslav Moiseyev)

The Bronze Age Okunevo culture is a unique phenomenon in the archeology of the southern and western Siberia, first of all due to its complex burial traditions and very rich art heritage that testify to the developed spiritual and religious views of the Okunevo people (Gass 2011). Although single kurgans and burials were excavated more than a hundred years ago (Savinov 2007; Vadetskaya 1986: 27, 28) the Okunevo culture was recognized and described as an independent cultural phenomenon only after excavations of Chernovaya 8 burial place by G.A. Makimenkov in 1962-1963 (Maksimenkov 1965, 1975, 1980). The culture was named after one of the earliest explored Okunevo burials in the Okunev ulus (Komarova 1947).

The Okunevo culture is represented mostly by burial grounds. Currently 62 Okunevo kurgans consisting of more than 500 burials and 60 single burials have been studied. Although several cases of the presence of Okunevo ceramics in cultural layers of multilayer settlements have been reported it is still not possible to connect traces of any buildings or hearths with this culture. Numeral engravings on rock “Pisanitsy” and stone stellas with complex drawings are the unique character of the Okunevo culture (Leont’ev et al. 2006).

All Okunevo sites have been found in the Minusinsk Basin which is located along the middle part of the Enisey River. This small territory of about 350 by 100 km is totally surrounded by the Eastern and Western Sayans mountains on one side and Kuznetsk Alatau on the other. Obviously such geographical isolation restricted population contacts of the Okunevo people with human groups in adjacent regions. Another geographical factor which added to the uniqueness of the Okunevo culture is rather complex landscape of the Minusinsk Basin which includes steppe, forest-steppe and taiga environments. This variation provided the opportunity to combine different models of economic activity the arrival of cattle breeding has been a principal source of discussion concerning origin of the Okunevo culture. Maksimentkov suggested that Okunevo culture was developed by the local Neolithic tribes of the Krasnoyarsk-Kansk forest-steppe who lived to the north of the Minusinsk Basin. After adopting cattle breeding and metal production from Afanasievo people these groups superseded Afanasievo tribes in the Minusinsk Basin (Maksimenkov, 1975: 36, 37). The second theory that is supported at the present time by most researchers suggests that Okunevo culture resulted from the interaction of local Neolithic hunter-gatherers with Western cattle breeders. This opinion is supported by evident parallels between early Okunevo burials and those of the Catacomb culture (Lazaretov 1995).

Based on results of excavations in the mid-1990s of a number of the Okunevo sites of the Uybat river basin, I. P. Lazaretov suggested dividing Okunevo culture on early Uybat and late Chernovaya periods (Lazaretov 1997). This was supported by most researchers. Later D. G. Savinov suggested additional final period of Okunevo culture called Razliv, which is represented by materials from three sites: Chernovaya XI, Razliv X, and Strelka (Savinov 2005). This suggestion remains disputable because of difficulties in differentiating of the artifacts and burial practices in the abovementioned sites from those of the Chernovaya period.

Radiocarbon AMS dating of 50 Okunevo samples are within 2600–1800 BCE (Polyakov 2017; Polykov and Svyatko 2009; Svyatko at al. 2009). According to these studies the Uybat period is dated as 2600–2300 BCE, Chernovaya as 2200–1900 BCE, and Razliv later than 1800 BCE.

The Okunevo culture shares some elements of its material culture including pottery with a number of local cultures from adjacent areas such as the Samus', Elunino, Karacol, and Krotovo cultures of western Siberia and Altai, the Kanay type burials of eastern Kazakhstan, and the Okunevo-like culture of Tuva. This makes it possible to view all of them as belonging to "the ring of related Okunevo-like cultures" (Molodin 2006; Savinov 1997; Stambulnik and Chugunov 2006). Nevertheless, there is currently no sound evidence of the common origins of all these cultures. Neither that there are similarities in their material cultures resulting from contacts of these peoples nor that there are broad time-specific characteristics of the area can be excluded. Few sites excavated on the upper Enysey in Tuva share elements of their material culture with Okunevo burials, but in spite of their geographical closeness to the Minusinsk Basin, the excavators of the site do not include them in the Okunevo culture in a strict sense (Stambulnik and Chugunov 2006).

According to studies of cranial morphology the Okunevo people resulted from admixture of Western Bronze Age migrants and local Neolithic tribes. It was reported that in the early Okunevo burials individuals displayed rather contrasting cranial morphology. Interestingly females demonstrated more Asian traits than males (Gromov 1997). Many Okunevo skulls have occipital-temporal deformation, which can result from cradle-boarding infants (Benevolenskaya and Gromov, 1997; Gromov 1998). The suggestion that Okunevo people and American Indians had common ancestors was based on the study of both cranial metric and nonmetric traits (Kozintsev et al. 1999) and was recently supported by genetic data (Allentoft et al. 2015).

3. Archaeological cultures of the Baikal region from the Late Mesolithic to the Bronze Age (A. W. Weber, V. I. Bazaliiskii, O. I. Goriunova)

The middle Holocene hunter-gatherer archaeology of the Baikal region in East Siberia has attracted the attention of Western scholarship from roughly the middle of the 20th century (Chard 1958; Michael 1958; Okladnikov 1959; Tolstoy 1958). The main reason for this attention was the availability of high-quality materials from habitation sites (camps) and cemeteries, the latter typically with large numbers of well-preserved human skeletal materials—a rarity among prehistoric hunter-gatherers worldwide and especially in the boreal zone. For example, Weber and Bettinger (2010) report 184 documented cemeteries with a total of 1,026 graves and 1,182 burials (individuals). However, these numbers have since increased somewhat due to continued fieldwork. More information about Baikal hunter-gatherer cemeteries can be found in a few recent reviews in English (Bazaliiskii 2003, 2010; Weber 1994, 1995; Weber and Bazaliiskii 1996; Weber et al. 2002) and Russian (Bazaliiskii 2005; Goriunova 1997; Kharinskii and Sosnovskaia 2000; Turkin and Kharinskii 2004).

Beginning in the late 1990s, these materials have become the subject of research by an international and multidisciplinary Baikal Archaeology Project (BAP) led by scholars from the University of Alberta, Canada, and Irkutsk State University, Russia (Weber et al. 2010). The project seeks a better understanding of the processes leading to the spatial and temporal variation in hunter-gatherer adaptive strategies, including the mechanisms of culture change. Comprehensive examination of human skeletal materials from the region’s cemeteries features prominently in this effort. While most of the bioarchaeological work has centered on the large cemeteries of Lokomotiv, Shamanka II, Ust-Ida I, Khuzir-Nuge XIV, and Kurma XI—all excavated over the course of the last 20–30 years. A number of other, frequently smaller collections, have been examined too, although with a narrower range of methods. This research continues to include as many additional materials from previous excavations from the entire Baikal region as are still available for examination.

Results of the chronological, archaeological, zooarchaeological, and bioarchaeological research conducted under the auspices of BAP have been presented in a large number of technical reports (Bronk Ramsey et al. 2014; Faccia et al. 2014, 2016; Haverkort et al. 2008; Katzenberg et al. 2008, 2009, 2012; Lieverse et al. 2007a, 2007b; 2008, 2009, 2011, 2014a, 2014b, 2015, 2016, 2017; Link 1999; Losey et al. 2008, 2011, 2012, 2013a, 2013b; Mooder et al. 2005, 2006; Moussa et al. 2016; Nomokonova et al. 2011, 2013, 2015; Osipov et al. 2016; Scharlotta et al. 2013, 2014, 2016, n.d.; Schulting et al. 2014, 2015; Shepard et al. 2016; Temple et al. 2014; Waters-Rist et al. 2010, 2011, 2014, 2016; Weber et al. 1998, 2011, 2013, 2016a, 2016b; White et al. n.d.), a few monographs (Weber et al. 2007, 2008, 2012) and several generalizing accounts (Lieverse et al. 2011; Losey and Nomokonova 2017; Weber 1995; Weber and Bettinger 2010; Weber and McKenzie 2003; Weber et al. 2002; Weber et al. 2010; 2011).

Our current views on the subject, summarized below, emphasize the multiple changes in the cultural patterns and recognize similarities between the Early Neolithic (EN) and Late Neolithic-Early Bronze Age cultures (LN-EBA) in addition to key differences, which were at the center of our attention earlier:

<u>Late Mesolithic</u>: incipient cemeteries, hunting, some fishing and sealing, small, dispersed, and mobile population, limited social differentiation.

<u>Early Neolithic</u>: cemeteries, hunting, fishing and sealing, large, unevenly distributed population, physical and physiological stress, differential mobility, substantial social differentiation.

<u>Middle Neolithic</u>: no cemeteries, hunting, some fishing and sealing, small,

dispersed, and mobile population, limited social differentiation.

Late Neolithic: cemeteries, hunting, fishing and sealing, larger and evenly distributed population genetically different from EN, moderate physical and physiological stress, moderate mobility and social differentiation.

Early Bronze Age: cemeteries, hunting, fishing and sealing, large and evenly distributed population genetically continuous with LN, moderate physical and physiological stress, moderate mobility and social differentiation.

With more results and insights becoming available, the following points summarize the most interesting aspects about the nature of the middle Holocene hunter-gatherer culture history and process in the Baikal region:

1. Much spatiotemporal variation existed in diet, subsistence, genetic structure, population size and distribution, number and size of cemeteries, health and activity patterns, mobility and migrations, mortuary protocols as well as socio-political differentiation between the micro-regions of the broader Baikal region.
2. The most intriguing aspect of this variation is that the EN hunter-gatherer system appears to be more complex and spatially variable than subsequent systems.
3. Lastly, the overall impression seems to be that change between these periods in the Baikal region was rapid rather than gradual.

Even with this much progress achieved, key issues related to the mechanism leading to the documented spatial variation in hunter-gatherer cultural patterns and temporal change in the Baikal region remain to be investigated further and understood better. Previous attempts to analyze mtDNA recovered from Baikal’s human skeletal remains have already provided useful insights about these matters (Mooder et al. 2005, 2006; Moussa 2016; Naumova et al. 1997; Naumova and Rychkov 1998), and it is the expectation that the much-improved methods of ancient DNA research can provide even more important insights now that encourage us to launch a new round of DNA studies on Baikal’s middle Holocene hunter-gatherers. Of particular interest are genetic connections with the outside world as well as the internal genetic structure, gene flow, marriage patterns, pathogen presence, and sex of osteologically indeterminable skeletons.

REFERENCES

Allentoft, M. E. et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522: 167–72. DOI: 10.1038/nature14507.

Anthony, D. W. 2007. *The Horse, the Wheel, and Language*. Princeton Univ. Press, Princeton.

Anthony, D. W. and D. R. Brown. 2011. The secondary products revolution, horse-riding, and mounted warfare. *J. World Prehist.* 24(2–3): 131–60.

Anthony, D. W. and D. Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist* 1.1: 199–219.

Bazaliiskii, V. I. 2003. The Neolithic of the Baikal region on the basis of mortuary materials. In: A.W. Weber and H. G. McKenzie (eds.), *Prehistoric Foragers of the Cis-Baikal, Siberia*. Edmonton: Canadian Circumpolar Institute Press, 37–50.

Bazaliiskii, V. I. 2005. K probleme kul'turno-khronologicheskikh osobennostei pogrebal'nykh komplesov epokhi pozdnego mezolita. In: A.V. Kharinskii (ed.), *Sotsiogenez v severnoi Azii*. Irkutsk: Irkutskii tekhnicheskii universitet, I: 35–42.

Bazaliiskii, V. I. 2010. Mesolithic and Neolithic mortuary complexes in the Baikal region of Siberia. In: A. W. Weber, M. A. Katzenberg, and T. Schurr (eds.), *Prehistoric Hunter-Gatherers of the Baikal Region, Siberia: Bioarchaeological Studies of Past Lifeways*. Philadelphia: University of Pennsylvania Museum of Archaeology and Anthropology, 51–86.

Bazaliiskiy, V. I. and N. A. Savelyev. 2003. The wolf of Baikal: the "Lokomotiv" early Neolithic cemetery in Siberia (Russia). *Antiquity* 77 20–30.

Benecke, N. and A. von den Driesch. 2003. Horse exploitation in the Kazakh steppes during the Eneolithic and Bronze Age. In: M. Levine, A. C. Renfrew, and K. Boyle (eds.), *Prehistoric Steppe Adaptation and the Horse* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 69–82.

Benevolenskaya, Y. D., and A. V. Gromov. 1997. The morphology of occipital-temporal area of Okunevo skulls. In: D. G. Savinov, M. L. Podol'skij (eds.)*Okunevsky Sbornik*. Saint Petersburg, Petro-RIF, 288–93.

Bronk Ramsey, C. et al. 2014. 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(2): 789–99.
Brown, D. and D. Anthony. 1998. Bit wear, horseback riding and the Botai site in Kazakhstan. *J. Arch. Sci.* 25(4): 331–47.
Chard, C. S. 1958. Outline of the prehistory of Siberia: Part 1. The pre-metal periods. *Southwestern J. of Anth.* 14(1): 1–33.
Faccia, K. et al. 2014. Bone quality in prehistoric, Cis-Baikal forager femora: A micro-CT analysis of cortical canal microstructure. *Am. J. Phys. Anth.* 154(4): 486–97.
Faccia, K. et al. 2016. Diffuse idiopathic skeletal hyperostosis (DISH) in a middle Holocene forager from Lake Baikal, Siberia: Potential causes and the effect on quality of life. *Quat. International* 405(B): 66–79.
Frachetti, M. D. 2008. *Pastoralist landscapes and social interaction in Bronze Age Eurasia*. Berkeley: University of California Press.
Gass, A. 2011. *Frühbronzezeit am mittleren Enisej. Gräberfelde der frühbronzeseitlichen Okunev-Kultur in Minusinsker Becken*. Bonn: Habelt Verlag.
Gaunitz, C. et al. 2018. Ancient genomes revisit the ancestry of domestic and Przewalski’s horses. *Science*. DOI: 10.1126/science.aao3297.
Goriunova, O. I. 1997. *Serovskie pogrebeniia Priol’khon’ia.* Novosibirsk, Institut arkheologii i etnografii SO RAN.
Gromov, A. V. 1997. Origin and biological affinities of the Okunevo people of the Minusa basin. In: D. G. Savinov, M. L. Podol'skij (eds.) *Okunevsky Sbornik*. Saint Petersburg, Petro-RIF, 301–45.
Gromov, A. V. 1997. Cranial nonmetric characteristics of Okunevo people. In: D. G. Savinov, M. L. Podol'skij (eds.) *Okunevsky Sbornik*. Saint Petersburg, Petro-RIF, 294–300.
Gromov, A. V. 1998. To the problem of artificial deformation of Okunevo skulls. In: D. G. Savinov, M. L. Podol'skij (eds.) *Sibir’ v panorama tysecheletiy*. Novosibirsk: Nauka, I: 147–56.
Haak, W. et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522 207–11.
Haverkort, C. et al. 2008. Hunter-gatherer mobility strategies and resource use based on strontium isotope (^{87}Sr/^{86}Sr) analysis: A case study from Middle Holocene Lake Baikal. *J. Arch. Sci.* 35: 1265–80.
Jones, M. et al. 2011. Food globalization in prehistory. *World Arch*. 43(4): 665–

75.

Katzenberg, M. A. et al. 2009. Paleodiet reconstruction of Early Bronze Age Siberians from the site of Khuzhir-Nuge XIV, Lake Baikal. *J. Arch. Sci.* 36: 663–74.

Katzenberg, M. A. et al. 2012. Prehistoric dietary adaptations among hunter-fisher-gatherers from the Little Sea of Lake Baikal, Siberia, Russian Federation. *J. Arch. Sci.* 39(8): 2612–26.

Katzenberg, M. A. and A. W. Weber. 1999. Stable isotope ecology and paleodiet in the Lake Baikal region of Siberia. *J. Arch. Sci.* 26(6): 651–59.

Kharinskii, A. V. and N. S. Sosnovskaia. *Baikal’skaia Sibir’ v drevnosti* 2(2): 66–100.

Kislenko, A. and N. Tatarintseva. 1999. The Eastern Ural steppe at the end of the Stone Age. In: M. Levine, Y. Rassamakin, A. Kislenko, and N. Tatarintseva, *Late Prehistoric Exploitation of the Eurasian Steppe* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 183–216.

Komarova, M. N. Burials of the Okunev Ulus. *Sovetskaya Archaeologia* 9: 47–60.

Kozintsev, A. G. et al. 1999. Collateral relatives of American Indians among the Bronze Age populations of Siberia. *Am. J. Phys. Anth*. 108(2): 193–204.

Kremenetski, C. V. et al. 1997. Postglacial development of Kazakhstan pine forests. *Geog. Phys. et Quat.* 51(3): 391–404.

Lazaretov, I. P. 1995. *Problemy izucheniya okunevskoy kul’tury*. Saint-Petersburg, Institute of the History of Material Culture, 14–16.

Lazaretov, I. P. 1997. The Okunevo burial places of Uibat river. In: D. G. Savinov, M. L. Podol'skij (eds.), *Okunevsky Sbornik*. Saint Petersburg: Petro-RIF, 19–64.

Leont’ev, N. V. et al. 2006. *Statues and Stella*'s *of the Okunevo culture*. Abakan: Khakasskoye knizhnoye izdatel’stvo, 236 c.

Levine, M. 2004. Exploring the Criteria for Early Horse Domestication. In: M. Jones (ed.), *Traces of Ancestry: studies in honour of Colin Renfrew* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 115–26.

Levine, M. and A. Kislenko. 2002. New Eneolithic and Early Bronze Age Radiocarbon Dates for North Kazakhstan and South Siberia. In: K. Boyle, A. C.

Renfrew, and M. Levine (eds.), *Ancient Interactions: East and West in Eurasia* (McDonald Institute Monographs). Cambridge: McDonald Institute for Archaeological Research, 131–34.

Lieverse, A. R. 2010. Health and behavior in mid-Holocene Cis-Baikal: biological indicators of adaptation and cultural change. In: A. W. Weber, M. A. Katzenberg, and T. Schurr (eds.), *Prehistoric Hunter-Gatherers of the Baikal Region, Siberia: Bioarchaeological Studies of Past Life Ways*. Philadelphia: Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, 135–73.

Lieverse, A. R. et al. 2007a. Dental health indicators of hunter-gatherer adaptation and cultural change in Siberia's Cis-Baikal. *Am. J. Phys. Anth.* 134, 323–39.

Lieverse, A. R. et al. 2007b. Osteoarthritis in Siberia’s Cis-Baikal: Skeletal indicators of hunter-gatherer adaptation and cultural change. *Am. J. Phys. Anth.* 132: 1–16.

Lieverse, A. R. et al. 2008. Pronounced bilateral asymmetry of the complete upper extremity: A case from the Early Neolithic Baikal (Siberia). *Int. J. Osteoarch.* 18: 219–39.

Lieverse, A. R. et al. 2009. Upper limb musculoskeletal stress markers among middle Holocene foragers of Siberia’s Cis-Baikal Region. *Am. J. Phys. Anth.* 138: 458–72.

Lieverse, A. R. et al. 2011. The bioarchaeology of habitual activity and dietary change in the Siberian middle Holocene. In: R. Pinhasi and J. Stock (eds.), *Human Bioarchaeology of the Transition to Agriculture*. Chicester: Wiley, 265–91.

Lieverse, A. R. et al. 2013. Lower limb activity in the Cis-Baikal: Entheseal changes among middle Holocene Siberian foragers. *Am. J. Phys. Anth.* 150(3): 421–32.

Lieverse, A. R. et al. 2014a. Paleopathological description and diagnosis of metastatic carcinoma in an Early Bronze Age (4588±34 cal. BP) forager from the Cis-Baikal region of Eastern Siberia. *PLOS One* 9(12): e113919.

Lieverse, A. R. et al. 2014b. Point taken: An unusual case of incisor agenesis and mandibular trauma in Early Bronze Age Siberia. *Int. J. Paleopath.* 6: 53–59.

Lieverse, A. R. et al. 2015. Death by twins: A remarkable case of dystocic childbirth in Early Neolithic Siberia. *Antiquity* 89: 25–38.

Lieverse, A. R. et al. 2016. Revisiting osteoarthritis in the Cis-Baikal: Understanding behavioral variability and adaptation among middle Holocene foragers. *Quat. International* 405(B): 160–71.

Lieverse, A. R. et al. 2017. Penetrating spinal injury: An extraordinary case of survival in Early Neolithic Siberia. *Int. J. Osteoarch.* 27: 508–14.

Link, D. W. 1999. Boreal forest hunter-gatherer demography and health during the middle Holocene of the Cis-Baikal, Siberia. *Arctic Anth.* 36(1–2): 51–72.

Losey, R. J. and T. Nomokonova, eds. 2016. *Holocene Zooarchaeology of Cis-Baikal. Archaeology in China and East Asia Vol 6, Northern Hunter-Gatherers Research Series Vol 7.* Berlin: German Archaeological Institute.

Losey, R. J. et al. 2008. Fishing ancient Lake Baikal, Siberia: Inferences from the reconstruction of harvested perch (*Perca fluviatilis*) size. *J. Arch. Sci.* 35: 577–90.

Losey, R. J. et al. 2011. Canids as persons: Early Neolithic dog and wolf burials, Cis-Baikal, Siberia. *J. Anth. Arch.* 30,: 174–89.

Losey, R. J. et al. 2012. Fish and fishing in Holocene Cis-Baikal, Siberia: A review. *J. Island and Coastal Arch.* 7: 126–45.

Losey, R. J. et al. 2013a. The bear-able likeness of being: Ursine remains at the Shamanka II cemetery, Lake Baikal, Siberia. In: C. Watts (ed.), *Relational*

Archaeologies. London: Routledge, 65–96.

Losey, R. J. et al. 2013b. Burying dogs in ancient Cis-Baikal, Siberia: Temporal trends and relationships with human diet and subsistence practices. *PLOS One* 8(5): e63740.

Ludwig, A. et al. 2009. Coat color variation at the beginning of horse domestication. *Science* 324(5926): 485–85.

Maksimenkov, G. A. 1965. *Okunevo culture of the Southern Siberia. Novoye v sovetskoy arkheologii (materially I issledovaniya v sovetskoy arkheologii N. 130).* Moscow: Nauka.

Maksimenkov, G. A. 1975. Thesis. Novosibirsk, Novosibirsk State University.

Maksimenkov, G. A. 1980. Chernovaya VIII – etalonnii pamjatnik okunevskoj kul’tury. In: J. B. Vadeckaja, N. V. Leont'ev, and G. A. Maksimenkov, (eds.), *Pamjatniki okunevskoj kul'tury*. Leningrad: Nauka, 3–26.

Merz, I. V. and V. K. Merz. 2013. Novye materialy rannego bronzovogo veka iz Zapadnoj chasti Kulundinskoj ravniny. In: G.A. Kubareva, , and V. P. Semibratov (eds.), *Sohranenie i izuchenie kul'turnogo nasledija Altajskogo kraja: materialy XVIII XIX regional'nyh nauchno-prakticheskih konferencij.* Barnaul: AZBUKA, 207–15.

Michael, H. N. 1958. The Neolithic Age in Eastern Siberia. *Trans. Am. Phil. Soc.* 48(2): 1–108.

Molodin, V. I. 2006. Karakol culture. In: *Okunevsky Sbornik 2: kultura i eyo okruzhenie*. Saint Petersburg: Elexis, 273–82.

Mooder, K. P. et al. 2005. Matrilineal affinities and prehistoric Siberian mortuary practices: A case study from Neolithic Lake Baikal. *J. Arch. Sci.* 32: 619–34.

Mooder, K. P. et al. 2006. Population affinities of Neolithic Siberians: A snapshot from prehistoric Lake Baikal. *Am. J. Phys. Anth.* 129: 349–61.

Moussa, N. M. et al. 2016, (in press). Y-Chromosomal DNA analyzed for four prehistoric cemeteries from Cis-Baikal, Siberia*. J. Arch. Sci. Reps.* dx.doi.org/10.1016/j.jasrep.2016.11.003.

Naumova, O. Y. et al. 1997. Molecular genetic characteristics of the Neolithic population of the Baikal region: RFLP of the ancient mtDNA from osseous remains found in the Ust’-Ida I burial ground. *Genetika* 33: 1418–25.

Naumova, O. Y. and S.Y. Rychkov. 1998. Siberian population of the New Stone Age: mtDNA haplotype diversity in the ancient population from the Ust’-Ida I burial ground, dated 4020–3210 BC by 14C. *Anthropologischer Anzeiger* 56:

1–6.
Nomokonova, T. et al. 2011. Use of Ulan-Khada bay at Lake Baikal in the Holocene (based on fauna remains). *Archaeology, Ethnology and Anthropology of Eurasia* 39(2): 30–36.
Nomokonova, T. et al. 2013. 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. Q*uat. International* 290–291: 110–125.
Nomokonova, T. et al. 2015. A 9,000 year history of seal hunting on Lake Baikal, Siberia: The zooarchaeology of Sagan-Zaba II. *PLoS One* 10(5): e0125314.
Okladnikov, A. 1974. *Neoliticheskie pamiatniki Angary*. Novosibirsk, Nauka.
Okladnikov, A. P. 1959. *Ancient populations of Siberia and its culture. Russian Translations of the Peabody Museum of Archaeology and Ethnology Vol 1.* Cambridge, MA: Harvard University.
Olsen, S. L. 2000. The secular and sacred roles of dogs at Botai, North Kazakhstan. In: S. Crockford (ed.), *Dogs Through Time: An Archaeological Perspective*. Oxford: BAR, 71–92.
Olsen, S. L. 2006. Early horse domestication: weighing the evidence. In: S. L. Olsen, S. Grant, A. M. Choyke, and L. Bartosiewicz (eds.), *Horses and Humans: The Evolution of Human-Equine Relations*. Oxford: Archaeopress, 81–114.
Olsen, S. L. et al. 2006. Community organization among Copper Age sedentary horse pastoralists of Kazakhstan. In: D. L. Peterson et al. (eds.) *Beyond the Steppe and the Sown*. Leiden: Brill, 89–111.
Osipov, B. et al. 2016. Evidence for genetic and behavioral adaptations in the ontogeny of prehistoric hunter-gatherer limb robusticity. *Quat. International* 405(B): 134–46.
Outram, A. K. Animal Domestications. In: V. Cummings, P. Jordan, and V. Hemingway (eds.), *Oxford Handbook of the Archaeology and Anthropology of Hunter Gatherers*. Oxford: OUP, 749–66.
Outram, A. K. et al. 2009. The earliest horse harnessing and milking. *Science* 323: 1332–35.
Ovchinnikov, M. P. 1904. *Materialy dlia izucheniia drevnostei v okrestnostiakh Irkutska. Izvestia VSORGO Vol XXXV.* Irkutsk: Irkutskaia tipolitografiia.

Polyakov, A. V. 2017. Radiocarbon dating of the Ocunevo culture. Zapiski Instituta istorii I material’noy kul’tury RAN. N 16.

Polyakov, A. V. and S. V. Svyatko. 2009. Radiocarbon dating of the Neolithic – Early Iron Age middle Enisey archaeological sites: results and new data. In: A. A. Tishkin (ed.), *Teoriya I praktika arkheologicheskikh issledovaniy*. Barnaul: Altayskiy Universitet, V: 20–56.

Sablin, M. V. and G. A. Khlopachev. 2002. The earliest Ice Age dogs: evidence from Eliseevichi. *Current Anthropology* 43: 795–99.

Savinov, D. G. 1997. The problems of the studies of the Okunevo culture. In: D. G. Savinov and M. L. Podol'skij (eds.), *Okunevsky Sbornik*. Saint Petersburg: Petro-RIF, 1–7.

Savinov, D. G. and M. L. Podol’skij. 1997. Introduction to Okunevo collections. In: D. G. Savinov and M. L. Podol'skij (eds.), *Okunevsky Sbornik*. Saint Petersburg: Petro-RIF, 7–18.

Savinov, D. G. 2005. To the problem of the definition of the latest period of the Okunevo culture. *Teoria I praktika arkheologicheskikh issledovaniy* 1: 28–34.

Savolainen, P. et al. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610–13.

Scharlotta, I. et al. 2016. Social consequences of increased reliance on fishing in middle Holocene Cis-Baikal: Relating fishing gear, axes, and social status at the Shamanka II cemetery, Lake Baikal, Siberia. *Quat. International* 419(C): 99–132.

Scharlotta, I. et al. 2013. Microsampling of human bones for mobility studies: diagenetic impacts and potentials for elemental isotopic research. *J. Arch. Sci.* 40: 4509–27.

Scharlotta, I. (submitted). “Mind the Gap” – Assessing methods for aligning age determination and growth rate in multi-molar sequences of dietary isotopic data. *Am. J. Phys. Anth.*

Scharlotta, I. and A. W. Weber. 2014. Mobility of middle Holocene foragers in the Cis-Baikal region, Siberia: Individual life history approach, strontium ratios, rare earth and trace elements. *Quat. International* 348: 37–65.

Schulting, R. J. et al. 2014. Freshwater reservoir offsets investigated through paired human–faunal ^{14}C dating and stable carbon and nitrogen isotope analysis at Lake Baikal, Siberia. *Radiocarbon* 56(3): 991–1008.

Schulting, R. J. et al. 2015. Highly variable freshwater reservoir offsets found along the Upper Lena watershed, Cis-Baikal, southern Siberia. *Radiocarbon* 57(4): 1–13.

Shepard, B. et al. 2016. Macro-regional interconnections among ancient hunter-gatherers of the Cis-Baikal, Eastern Siberia (Russia). *Quat. International* 419(C): 140–58.

Sherratt, A. 2006. The Trans-Eurasian exchange: the prehistory of Chinese relations with the West. In: V. Mair (ed.), *Contact and Exchange in the Ancient World*. Honolulu: Hawaii UP, 1–9.

Svyatko S. V. et al. 2009. New radiocarbon dates and a review of the chronology of prehistoric populations from the Minusinsk basin, Southern Siberia, Russia. *Radiocarbon* 51(1): 243–73.

Svyatko, S. V. et al. (in press). A lack of freshwater reservoir effects in human radiocarbon dates in the Eneolithic to Iron Age in the Minusinsk Basin. *Arch. and Anthropol. Sci.* doi:10.1007/s12520-016-0383-3 [http://link.springer.com/article/10.1007/s12520-016-0383-3/fulltext.html].

Vadetskaya, E. B. 1986. *The archaeological sites of the steppes of the middle Enisey*. Leningrad: Nauka.

Temple, D. H. et al. 2014. Skeletal growth in Early and Middle Holocene foragers from the Cis-Baikal region of East Siberia. *Am. J. Phys. Anth.* 153: 377–86.

Tiutrin, A. A. and V. I. Bazaliiskii. 1996. Mogil’nik v ust’e reki Idy v doline Angary. *Akheologiia i etnologiia Sibiri i Dal’nego Vostoka* 1: 85–90.

Tolstoy, P. 1958. Archaeology of the Lena basin and its new world relationships, Part I. *American Antiquity* 23: 397–418.

Turkin, G. V. and A. V. Kharinskii. 2004. Mogil’nik Shamanka II: K voprosu o khronologii i kul’turnoi prinadlezhnosti pogrebal’nykh kompleksov neolita–bronzovogo veka na iuzhnom Baikale. *Izvestiia Laboratorii drevnikh tekhnologii Irkutskogo gosudarstvennogo tekhnicheskogo universiteta* 2: 124–58.

Waters-Rist, A. L. 2012. A unique case of mandibular osteomyelitis arising from tooth germ infection in a 7000 year old infant from Siberia. *Dental Anthropology* 25: 15–25.

Waters-Rist, A. L. et al. 2010. Activity–induced dental modification in Holocene Siberian hunter-fisher-gatherers. *Am. J. Phys. Anth.* 20(2): 266–78.

Waters-Rist, A. L. et al. 2011. Infant and child diet in Neolithic hunter-fisher-gatherers from Cis-Baikal, Siberia: Intra-long bone stable nitrogen and carbon isotope ratios. *Am. J. Phys. Anth.* 146: 225–41.

Waters-Rist, A. L. et al. 2014. Multicomponent analyses of a hydatid cyst from an Early Neolithic hunter-fisher-gatherer from Lake Baikal, Siberia. *J. Arch. Sci.* 50: 51–62.

Waters-Rist, A. L. et al. 2016. Evaluating the biological discontinuity hypothesis of Cis-Baikal Early versus Late Neolithic/Bronze Age populations using dental non-metric traits. *Quat. International* 405(B): 122–33.

Wayne, R. K. et al. 2006. Genetic analysis of dog domestication. In: M. A. Zeder et al. (eds.), *Documenting domestication: new genetic and archaeological paradigms*. Berkeley: University of California Press, Berkeley, 279–93.

Weber, A. W. 1994. Social evolution among Neolithic and Early Bronze Age foragers in the Lake Baikal Region: New light on old models. *Arctic Anth.*

31(2): 1–15.

Weber, A. W. 1995. The Neolithic and Early Bronze Age of the Lake Baikal region, Siberia: A review of recent research. *J. World Prehist.* 9: 99–165.

Weber, A. W. and V. I. Bazaliiskii. 1996. Mortuary practices and social relations among the Neolithic foragers of the Angara and Lake Baikal Region: Retrospection and prospection. In: D. A. Meyer, P. C. Dawson, D. T. Hanna, (eds.), *Debating Complexity: Proceedings of the 26th Annual Chacmool Conference*. Calgary: Archaeological Association of the University of Calgary, 97–103.

Weber, A. W. and R. L. Bettinger. 2010. Middle Holocene hunter-gatherers of Cis-Baikal, Siberia: An overview for the new century. *J. of Anth. Arch.* 29: 491–506.

Weber, A. W. and O. I. Goriunova. 2013. Hunter-gatherer migrations, mobility and social relations: A case study from the Bronze Age Baikal region, Siberia. *J. of Anth. Arch.* 32(6): 330–46.

Weber, A. W. et al. 1998. Patterns of prehistoric procurement of seal at Lake Baikal: A zooarchaeological contribution to the study of past foraging economies in Siberia. *J. Arch. Sci.* 25: 215–27.

Weber, A. W. et al. 2002. Hunter-gatherer culture change and continuity in the Middle Holocene Cis-Baikal, Siberia. *J. of Anth. Arch.* 21: 230–299.

Weber, A. W. et al., eds. 2007. *KHUZHIR-NUGE XIV, a Middle Holocene Hunter-gatherer Cemetery on Lake Baikal, Siberia: Osteological Materials. Northern Hunter-Gatherers Research Series Vol 3.* Edmonton: Canadian Circumpolar Institute Press.

Weber, A. W. et al., eds. 2008. *KHUZHIR-NUGE XIV, a Middle Holocene Hunter-gatherer Cemetery on Lake Baikal, Siberia: Archaeological Materials. Northern Hunter-Gatherers Research Series Vol 4.* Edmonton: Canadian Circumpolar Institute Press.

Weber, A. W. et al., eds. 2010. *Prehistoric Hunter-Gatherers of the Baikal Region, Siberia: Bioarchaeological Studies of Past Lifeways*. Philadelphia: University of Pennsylvania Museum Press.

Weber, A. W. et al. 2011. Hunter-gatherer foraging ranges, migrations, and travel in the middle Holocene Baikal region of Siberia: Insights from carbon and nitrogen stable isotope signatures. *J. of Anth. Arch.* 30(4): 523–48.

Weber, A. W. et al., eds. 2012. *KURMA XI, a Middle Holocene hunter-gatherer cemetery on Lake Baikal, Siberia: Archaeological and Osteological Materials. Archaeology in China and East Asia Vol 3, Northern Hunter-Gatherers*

Research Series Vol 6. Berlin: German Archaeological Institute.

Weber, A. W. et al. 2013. Environmental change and cultural dynamics of Holocene hunter-gatherers in Northeast Asia: Comparative analyses and research potentials in Cis-Baikal (Siberia) and Hokkaido (Japan). *Quat. International* 290–291: 3–20.

Weber, A.W. et al. 2016a. Biogeochemical data from the Shamanka II Early Neolithic cemetery on southwest Baikal: Chronological and dietary patterns. *Quat. International* 405(B): 233–54.
Weber, A. W. et al. 2016b. Chronology of middle Holocene hunter-gatherers in the Cis-Baikal region of Siberia: Corrections based on examination of the freshwater reservoir effect. *Quat. International* 419(C): 74–98.
White, J. A. et al. (submitted). Integrated stable isotopic and radiocarbon analyses of Neolithic and Bronze Age hunter-gatherers from Lake Baikal's Little Sea and Upper Lena River Regions. *J. Arch. Sci.*
Zaibert, V. F. 2009. *Botaiskaya Kultura*. Almaty: KazAkparat.
Zeder, M. A. 2012. The domestication of animals. *J. Anthropol. Res.* 68(2), 161-190 (2012).
Zaibert, V. F. et al. 2007. *Tajny Drevnyej Stepi: Issledovaniya Poseleniya Botai (2004-2006)*. Kokshetau: Kokshetau University.

Linguistic supplement to Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian

AUTHORS

Guus Kroonen[1,3], Gojko Barjamovic[2], and Michaël Peyrot[3].

AFFILIATIONS

[1]Department of Nordic Studies and Linguistics, University of Copenhagen, Denmark.

[2]Department of Near Eastern Languages and Civilizations, Harvard University, USA.

[3]Leiden University Centre for Linguistics, Leiden University, The Netherlands.

ABSTRACT

We recount the evidence for the so-called "Steppe Hypothesis" discussed in Damgaard et al. 2018 and offer a revised linguistic and historical model for the prehistoric dispersal of three important Indo-European language subgroups—the Anatolian Indo-European languages into Anatolia, the Tocharian languages into Inner Asia, and the Indo-Iranian languages into South Asia—based on the newly analysed archaeogenetic data.

1. Origins and dispersals of the Indo-European languages

The Indo-European language family is among the largest in the world and is spoken by ca. 44% of the global population (Simons and Fennig 2017). It derives from a prehistoric and extinct dialect continuum spoken in an area that can be approximated only by the combined study of historical linguistics, archaeology, and ancient human population genetics. From this hypothetical nucleus, the Indo-European parent language, also known as Proto-Indo-European, split into a variety of subgroups that dispersed over large distances in prehistoric times. At their earliest attestations, the branches Italic, Celtic, Germanic, Balto-Slavic, Albanian, Greek, Anatolian, Armenian, Indo-Iranian, and Tocharian already covered a large area across Eurasia, stretching from Atlantic Europe in the West to the Taklamakan Desert of China in the East.

The time and location of the Proto-Indo-European linguistic unity is uncertain, since it long predates the earliest historical records. A *terminus ante quem* for the dissolution of Proto-Indo-European is offered by the earliest appearances of the individual daughter languages, e.g. Mycenaean Greek in the 16th century BCE, Indo-Aryan in North Syrian texts from the 18th century BCE, and Anatolian as early as the 25th century BCE. Concerning the deeper origin of the proto-language, various theories exist (cf. e.g. Gamkrelidze and Ivanov 1995; Renfrew 1987, 1999). Here we focus on the prevalent “Steppe Hypothesis,” which places the speakers of Proto-Indo-European on the Pontic steppe in the 4th millennium BCE (Anthony 1995, 2007; Gimbutas 1965; Mallory 1989).

The time and location postulated by proponents of this hypothesis are dictated by cultural markers contained in the Proto-Indo-European vocabulary itself. These markers, which are found in the reconstructed lexicon shared by various Indo-European subgroups, consist of archaeologically salient terminology related to 1) copper-based metallurgy, 2) pastoral nomadism, 3) horse domestication (see Outram et al. 2018), 4) wheeled vehicles, and 5) wool production (e.g. Beekes 2011; Mallory and Adams 1997). Based on this reconstructed cultural assemblage, Proto-Indo-European linguistic unity must approximately be placed in the Chalcolithic (Copper Age) and at a location where the social order and technologies found in the shared vocabulary were extant.

Although material culture and linguistic entities do not generally match, archaeological and linguistic reconstructions of prehistory can be compared to see where and when they might overlap. The area covered by the archaeological Yamnaya horizon of the Pontic steppes 3000–2400 BCE has long been held as a suitable candidate for a region from where speakers of Proto-Indo-European (sometimes excluding the Anatolian branch) could have dispersed (Anthony 2007; Chang et al. 2015; Gimbutas 1965; Mallory 1989). Expansions of Yamnaya material culture into Europe (Corded Ware culture; Kristiansen et al. 2017) and southern Siberia (Afanasievo culture) have recently been documented through studies, including the present one, of the archaeological and genetic data, which suggest that such expansions were at least partly linked with a movement of Yamnaya culture-bearing populations (Allentoft 2015; Haak 2012). This supports their potential as vectors for the spread of Indo-European languages to the areas where they are first attested.

2. The Anatolian Indo-European languages

The Anatolian branch is an extinct subclade of the Indo-European language family attested from the 25th century BCE onwards (see below) that consists of Hittite (known 20th–12th centuries BCE), Luwian (known 20th–7th centuries BCE), and a number of less well-attested members, such as Carian, Lycian, Lydian, and Palaic. Hittite is mainly attested through thousands of clay tablets inscribed in cuneiform writing obtained from the institutional archives of the Hittite state (ca. 1650–1180 BCE).

The position of the Anatolian branch within the Indo-European family tree is still debated (cf. Melchert fthc.). Although Hittite is closely related to the other Indo-European languages, it features some divergent characteristics, such as 1) a retention of linguistic archaisms, 2) uniquely Anatolian innovations, and 3) an absence of innovations found in languages of the other branches.

After the identification of Hittite as an Indo-European language (Knudtzon 1902) and its decipherment (Hrozný 1915), these divergent characteristics prompted the view that Anatolian split off from Proto-Indo-European earlier than the other branches. This gave rise to the so-called Indo-Anatolian (or Indo-Hittite) Hypothesis (Sturtevant 1933: 30), whose proponents claimed that Anatolian descended from a sister language of Proto-Indo-European, rather than being a daughter. The two would thus have derived from an older common ancestor. While gaining traction in the latter half of the 20th century, the Indo-Anatolian Hypothesis recently lost acceptance following attempts to remodel the reconstruction of Proto-Indo-European after the Anatolian branch (cf. Adrados 2007; Kuryłowicz 1964; Watkins 1969) and a lack of consensus concerning identification of the putative Anatolian archaisms (see esp. Rieken 2009). While the vast majority of Indo-Europeanists would still agree that Anatolian is the most likely branch to have split off first (cf. Lehrman 1998; Melchert 2017: 194; Melchert fthc.: 52–53), and evidence in support of the Indo-Anatolian Hypothesis is mounting (cf. Kloekhorst 2016), the view that Anatolian is a sister rather than a daughter language of Proto-Indo-European remains disputed (Melchert 2017: 194).

2.1 Native Sources and Terminologies

The term “Hittite” in current terminology is not cognate with ancient usage. The state itself was known to its contemporaries as “The Land of Hat(t)i” (del Monte and Tischler 1978: 101)—a non-declinable noun of uncertain origin (Weeden 2011: 247)—while the language that we in modern time refer to as “Hittite” was known to its speakers as *neš(umn)ili*, i.e. the language of Neša or Kaneš, the modern-day site of Kültepe near Kayseri.

Some 23,000 inscribed clay tablets have been unearthed at Kaneš (Larsen 2015), but these belong to a period (ca. 1920–1720 BCE) before the first texts were written in the Hittite language. Instead, they constitute a body of records kept by an Assyrian merchant community who settled at the site and wrote in their own Semitic language, the Old Assyrian dialect of Akkadian. The records make frequent reference to the local Anatolian population, which was multilingual and took part in a larger sphere of close commercial exchange (see Kristiansen et al. 2018). They also record hundreds of personal names belonging to individuals settled in the region of Kaneš that can be related to various languages, including Hittite, Luwian, Hurrian, and Hattian (Laroche 1966, 1981; Wilhelm 2008; Zehnder 2010). Finally, the merchant records contain a number of Anatolian Indo-European loanwords (Bilgiç 1954; Dercksen 2007; Schwemer 2005–2006: 221–224) adopted by the Assyrian community.

However, the Assyrian merchants made no distinction between local groups along ethnic or linguistic lines and applied the blanket term *nu(w)ā’um* to refer to the Anatolian population at large (Goedegebuure 2008 with references). Instead, they distinguished individuals according to statehood (e.g. “the man from Wašhaniya,” “the Kanišite”), and used terms, such as “the Land” (*ša mātim / libbi mātim*) to refer to Anatolia or its heartland (Barjamovic 2011). Alongside the general impression of Kaneš as a cosmopolitan society characterized by hybrid artistic and religious traditions (Larsen and Lassen 2012), the records from Kaneš show a highly mixed linguistic milieu with usage apparently linked to context (trade languages, ritual languages, etc.) in which language did not serve as an ethnic marker.

2.2 Geographical origins and spread of the Anatolian Indo-European languages

The prehistory of the Anatolian Indo-European branch remains poorly understood. There is general consensus among Hittitologists that it constitutes an intrusive branch (Melchert 2003: 23), the dispersal of the Indo-European languages commonly being linked to the Yamnaya archaeological and genetic expansions from the Pontic-Caspian steppe (Allentoft 2015; Anthony 2007; Mallory 1989). It clearly did not evolve *in situ* from a local source (Bouckaert et al. 2012; Renfrew 1987), but a lack of concrete archaeological or genetic evidence for an influx of outside groups means that any exact timing or route of migration of Anatolian Indo-European speakers to Anatolia is debated. Some scholars have suggested that the split of Proto-Anatolian may have been early enough to have happened outside Anatolia, implying several movements of Anatolian-speaking groups (Steiner 1990: 202f.). Without any trace of Anatolian languages outside Anatolia, however, the default hypothesis remains that Proto-Anatolian split up into different dialects in Anatolia itself, probably sometime in the mid- to late 4th millennium BCE.

Despite a general agreement on a Pontic-Caspian origin of the Anatolian Indo-European language family, it is currently impossible to determine on linguistic grounds whether the language reached Anatolia through the Balkans in the West (Anthony 2007; Mallory 1989: 30; Melchert 2003; Steiner 1990; Watkins 2006: 50) or through the Caucasus in the East (Kristiansen 2005: 77; Stefanini 2002; Winn 1981). From their earliest attestations, the Anatolian languages are clustered in Anatolia, and if the distribution reflects a prehistoric linguistic speciation event (as argued by Oettinger 2002: 52), then it may be taken as an indication that the arrival and disintegration of Proto-Anatolian language took place in the same area (Steiner 1981: 169). However, others have reasoned that the estimated period between the dissolution of the Proto-Anatolian language and the attestation of the individual daughter languages is extensive enough to allow for prehistoric mobility within Anatolia, theoretically leaving plenty of time for secondary East-to-West dispersals (cf. Melchert 2003: 25).

Whatever the case may be, there are no linguistic indications for any mass migration of steppe-derived Anatolian speakers dominating or replacing local populations. Rather, the Anatolian Indo-European languages appear in history as an organically integrated part of the linguistic landscape. In lexicon, syntax, and phonology, the second millennium languages of Anatolia formed a convergent, diffusional linguistic area (Watkins 2001: 54). Though the presence of an Indo-European language itself demonstrates that a certain number of speakers must have entered the area, the establishment of the Anatolian Indo-European branch in Anatolia is likely to have happened through a long-term process of infiltration and acculturalization rather than through mass immigration or elite dominance (Melchert 2003: 25).

Furthermore, the genetic results presented in Damgaard et al. 2018 show no indication of a large-scale intrusion of a steppe population. The EHG ancestry detected in individuals associated with both Yamnaya (3000–2400 BCE) and the Maykop culture (3700–3000 BCE) (in prep.) is absent from our Anatolian specimens, suggesting that neither archaeological horizon constitutes a suitable candidate for a “homeland” or “stepping stone” for the origin or spread of Anatolian Indo-European speakers to Anatolia. However, with the archaeological and genetic data presented here, we cannot reject a continuous small-scale influx of mixed groups from the direction of the Caucasus during the Chalcolithic period of the 4th millennium BCE.

2.3 Dating Anatolian Indo-European – Evidence from Ebla

We stress that the presence of the Anatolian Indo-European language in Anatolia must be much older than the first cuneiform evidence. Anatolian personal names resembling those appearing in the Assyrian trade records are attested approximately half a millennium earlier among individuals said to be from the state of Armi. These are recorded in texts found in the palatial archives of the city of Ebla in Syria, dated to the 25–24th centuries BCE (Bonechi 1990).

The location of Armi remains unknown and is debated (Archi 2011; Bonechi 2016). It was clearly a state with multiple urban centres and was in a position to control Ebla's access to commodities that can be securely associated with the Anatolian highlands, chiefly metal. Among the individuals listed as coming from Armi, some bear names of unknown derivation while others may have had names that are Semitic in origin. It is not always clear whether the latter are in fact merely the names of Eblaites active in Armi (Winters in prep.).

However, a small group of ca. twenty names connected to Armi build on what appear to be well-known Anatolian roots and endings, such as *-(w)anda/u*, *-(w)aššu*, *-tala*, and *-ili/u*, cf. *A-la-lu-wa-du*, *A-li-lu-wa-da*, *A-li-wa-da*, *A-li-wa-du*, *A-lu-wa-da*, *A-lu-wa-du*, *Ar-zi-tá-la*, *Ba-mi-a-du*, *Ba-wi-a-du*, *Du-du-wa-šu*, *Ha-áš-ti-lu*, *Hu-da-šu*, *Mi-mi-a-du*, *Mu-lu-wa-du*, *Tar*$_5$*-hi-li*, and *Ù-la-ma-du* (Archi 2011: 21–25; Bonechi 1990). The Eblaite script does not always distinguish voiced and voiceless consonants and ignores germinates (Catagnoti 2012). This renders it difficult to establish an exact reading of the names and makes it impossible at present to determine the language or languages to which the names from Armi belong with any certainty, except to say that they clearly fall within the Anatolian Indo-European family.

Regardless of their exact linguistic background, however, the implications held by the presence of individuals with identifiable Anatolian

Indo-European names in Southern Turkey at this early point in history for the development of Indo-European languages and the Anatolian split are significant. The dissolution of Proto-Anatolian into its daughter languages is usually estimated by linguists to have taken place at least several centuries (Melchert 2003: 23), if not more than a millennium (Anthony 2007: 46; Steiner 1990: 204), before the start of the written record. With the retrojection of Anatolian Indo-European speakers in Anatolia by approximately 500 years, the period of Proto-Anatolian linguistic unity can be pushed further back in time.

Also, since the onomastic evidence from Armi is contemporaneous with the Yamnaya culture (3000–2400 BCE), a scenario in which the Anatolian Indo-European language was linguistically derived from Indo-European speakers originating in this culture can be rejected. This important result offers new support for the Indo-Hittite Hypothesis (see above) and strengthens the case for an Indo-Hittite-speaking ancestral population from which both Proto-Anatolian and residual Proto-Indo-European split off no later than the 4th millennium BCE.

3. Inner Asia: the Tocharian languages

The only known branch of the Indo-European language family thought to have been spoken in Inner Asia prior to the Bronze Age is represented by the two closely related languages Tocharian A and Tocharian B. These are attested through Buddhist manuscripts found in the Tarim Basin in Northwest China dating from ca. 500–1000 CE. On their way to the Tarim Basin, the linguistic ancestors of the speakers of Tocharian must at some point have crossed the Eurasian steppe from the region of origin of the Indo-European language family. It is usually assumed that the Afanasievo culture of the Altai region (ca. 3000–2500 BCE; cf. Vadeckaja, Poljakov, and Stepanova 2014) represents an early, intermediate phase in their prehistory (Anthony 2007: 264–265; Mallory 1989: 62–63).

An obvious difficulty with this identification is that the language or languages spoken by people associated with a prehistoric archaeological culture are unknown. It is theoretically possible that the cultural remains which we identify as Afanasievo were associated with speakers of multiple languages, or with speakers of an Indo-European language that was not ancestral to Tocharian and left no trace in the written record. Another issue is the archaeological problem of linking the Afanasievo culture to the historical Tocharian speakers across a time gap of ca. 3000 years.

An intermediate stage has been sought in the oldest so-called Tarim Mummies, which date to ca. 1800 BCE (Mallory and Mair 2000; Wáng 1999). However, also the language(s) spoken by the people(s) who buried the Tarim Mummies remain unknown, and any connection between them and the Afanasievo culture on the one hand or the historical speakers of Tocharian on the other has yet to be demonstrated (cf. also Mallory 2015; Peyrot 2017).

In spite of these evident problems, the identification of the Afanasievo culture with the ancestors of the speakers of Tocharian currently provides the best explanation for the evidence at hand. This identification is founded upon a series of considerations. First, despite their geographical proximity, the ancestors of the speakers of Tocharian cannot be associated with the Indo-Iranian Sintashta and Andronovo cultures (discussed below), since Tocharian is not more closely affiliated with Indo-Iranian than with any other branch of Indo-European. While the Indo-Iranian languages belong to the so-called *satəm* languages, as seen e.g. by Vedic *śatám* (hundred) and Avestan *satəm* itself, Tocharian belongs to the *centum* group, as shown by Tocharian B *kante*, A *känt* (hundred). The fact that Tocharian is so different from the Indo-Iranian languages can only be explained by assuming an extensive period of linguistic separation. Second, the Afanasievo culture could be a good match chronologically, seeing as it precedes the spread of the Andronovo culture in the Eurasian steppe (see below). The latter is likely to have been Iranian-speaking (or perhaps in part Indo-Iranian-speaking) and an identification of the ancestral Tocharian speakers with the Afanasievo culture leaves time for them to cross the Eurasian steppe without coming into linguistic contact with the Iranian or Indo-Iranian speakers who dominated the steppe region in the Bronze Age and Iron Age. Third, although the core area of the Afanasievo culture is located in the northern Altai, about 1000 km north of the Tarim Basin, it is situated on roughly the same eastern longitude as the later Tocharian sites, and is therefore geographically a relatively appropriate match. Fourth, Afanasievo material culture is generally said to be closely related to the Yamnaya (Anthony 2007: 307–311; Chernykh 1992: 28; Vadeckaja 1986: 22), and individuals attributed to these cultures show closely related genetic ancestry (Allentoft et al. 2015). The Yamnaya culture is widely acknowledged to have driven, for a large part, the spread of the Indo-European languages into Europe, and Afanasievo may therefore have had a comparable linguistic impact in Asia.

In Damgaard et al. 2018, we present a high-coverage genome from Karagash that is consistent with previously published Yamnaya and Afanasievo genomes. This may hold implications for a better understanding of the between Yamnaya and Afanasievo, as it identifies related individuals in the area that separates the two cultures (3,000–4,000 km distant from one another) and provides further evidence for a possible route connecting them (Anthony 2007: 309; Mallory 1989: 225–226).

Further, we observe that there is no close genetic relationship between the Botai individuals and the Yamnaya or Afanasievo profiles (Damgaard et al. 2018). The language(s) of the people associated with the Botai culture is unknown, so we cannot link this finding to any linguistic observation, but simply note that there is no evidence that an early stage of Tocharian was impacted by any language of horse herders such as the Botai. For instance, Tocharian has inherited the word for “horse” from Proto-Indo-European, i.e. Tocharian B *yakwe* and Tocharian A *yuk*, both going back to PIE *h_1eḱuo-. Hardly any technical terms related to horses or horse herding are attested in

Tocharian, but there is no reason at present to assume a strong influence from a language of horse herders. This is consistent with the apparent lack of a genetic flow between the Botai samples and those associated with Yamnaya and Afanasievo.

Finally, we find that two of the individuals analysed are genetically almost indistinguishable from specimens associated with the Okunëvo culture even though they were buried in Afanasievo-like pits, and that 19 Okunëvo samples are found to have been admixed with 10–20% Yamnaya/Afanasievo ancestry (Damgaard et al. 2018). The appearance of the Okunëvo culture (ca. 2500–2000 BCE) in the Altai region marks the end of the Afanasievo culture and may have caused members of the earlier population to leave the area and move south into the Tarim Basin. But our findings identify both a cultural overlap and genetic admixture between individuals associated with the Afanasievo and Okunëvo cultures, suggesting that the transition from one to the other was not necessarily abrupt and may have involved gradual processes of mutual acculturalization (see Outram et al. 2018). Future research may show whether any genetic ancestry from individuals associated with the Okunëvo culture was carried by descendants of those associated with the Afanasievo culture who supposedly moved south into the Tarim Basin. It is conceivable, for instance, that those who remained in the Altai region produced the mixed culture and ancestry after those descendants had left. In that case, no cultural, genetic or linguistic influence of populations associated with the Okunëvo culture would be expected in Tocharian speakers.

4. The Indo-Iranian languages

The Indo-Iranian languages form the dominant branch of Indo-European in Asia in terms of its wide distribution and large number of speakers. The branch is commonly divided into three main subgroups: Indo-Aryan (or Indic), Iranian, and the smaller group of Nuristani languages found on the border of Afghanistan and Pakistan, which occupy a dialectically intermediate position (Fussman 1972: 390; Morgenstierne 1973; Strand 1973). Indo-Aryan is most famously represented by Vedic Sanskrit, the language of the religious hymns of the Rig Veda. Iranian languages are attested from the 8th century BCE, the most important members being Old Persian, the language of the Achaemenid state elite, and Avestan, the sacred language of Zoroastrianism. Being spread over a large area, the Indo-Iranian languages and peoples had enormous impact on the linguistic and cultural landscape of Asia: Indo-Aryan (or Indic) with Hindi, Urdu, Bengali, and Punjabi as prominent modern representatives, and Iranian with widely spoken idioms, such as Farsi (Persian), Pashto, and Kurdish.

4.1 Dating the Indo-Iranian unity and split

Under the “Steppe Hypothesis,” the Indo-Iranian languages are not seen as indigenous to South Asia but rather as an intrusive branch from the northern steppe zone (cf. Anthony 2007: 408–411; Mallory 1989: 35–56; Parpola 1995; Witzel 1999, 2001). Important clues to the original location and dispersal of the Indo-Iranians into South and Southwest Asia are provided by the Indo-Iranian languages themselves.

The Indo-Aryan and Iranian languages share a common set of etymologically related terms related to equestrianism and chariotry (Malandra 1991). Since it can be shown that this terminology was inherited from their Proto-Indo-Iranian ancestor, rather than independently borrowed from a third language, the split of this ancestor into Indo-Aryan and Iranian languages must postdate these technological innovations. The earliest available archaeological evidence of two-wheeled chariots is dated to approximately 2000 BCE (Anthony 1995; Anthony and Ringe 2015; Kuznetsov 2006: 638–645; Teufer 2012: 282). This offers the earliest possible date so far for the end of Proto-Indo-Iranian as a linguistic unity. The reference to a *mariannu* in a text from Tell Leilān in Syria discussed below pushes the latest possible period of Indo-Iranian linguistic unity to the 18th century BCE.

The *terminus ante quem* for the disintegration of Proto-Indo-Iranian is provided by traces of early Indo-Aryan speakers in Southwest Asia. The text in Hittite CTH 284 dating to the 15th–14th centuries BCE gives detailed instructions by “Kikkuli, master horse trainer of the land of Mitanni.” It makes use of Indo-Iranian, or possibly Indo-Aryan terminology, including *wa-ša-an-na-* (training area), and *a-i-ka-*, *ti-e-ra-*, *pa-an-za-*, *ša-at-ta-*, *na-a-wa-ar-tan-na-* (one, three, five, seven, nine rounds). It is generally thought that this terminology was particularly linked to the Mitanni state (16th–14th centuries BCE), where names of Indo-Aryan derivation appear among the ruling class of a mostly Hurrian-speaking population (Mayrhofer 1982; Thieme 1960; Witzel 2001: 53–55). Indo-Aryan adjectives denoting horse colors are known from the texts of the provincial town of Nuzi on the eastern frontier of Mitanni, including *pabru-nnu-* (reddish brown), *parita-nnu-* (gray), *pinkara-nnu-* (reddish brown) (Mayrhofer 1966: 19, 1974: 15f., 1982: 76). Furthermore, “the Mitra-gods, the Varuna-gods, Indra, and the Nāsatya-gods” are listed among the divine witnesses of Mitanni in the treaty CTH 51 between its ruler Šatiwazza and Šuppiluliumas of the Land of Hatti (Beckman 1996: 43).

A recently discovered reference to *mariannu* in a letter from Tell Leilān in Northern Syria dating shortly before the end of Zimri-Lim’s reign in 1761 BCE (Eidem 2014: 142) extends the Indo-Aryan linguistic presence in Syria back two centuries prior to the formation of the Mitanni state. The word is generally seen as a Hurrianized form of the Indo-Aryan word **marya-* (man/youth) (von Dassow 2008: 96–97 with literature) and taken to refer to a type of military personnel associated with chariot warfare across the Near East (*eadem* pp. 268–314).

A debate on how to interpret the occurrence of these Indo-Aryan technical terms, divinities, and personal names in the Bronze Age state of Mitanni has gone on for more than a century (Winckler 1910: 291). Van Koppen 2017 has recently drawn attention to the near-contemporaneous appearance of a Kassite-speaking population in Babylonia as a possible model also for the Mitanni linguistic diffusion. From a linguistically heterogeneous migrant population coming from the Zagros, the Kassite group rose to power in Babylon, and its language and names as markers of identity became normative for their dynastic successors (*idem* p. 81).

The personal names with apparent Indo-Aryan etymologies persisted across a surprisingly large territory and appear as far apart as Nuzi in the east and Palestine in the west (Ramon 2016). Unlike the military and hippological terms, which were part of a technical vocabulary and adopted into local languages, the distinct naming practice and the list of divine witnesses appearing in the Šatiwazza treaty imply that elements that we define as Indo-Aryan played a role in maintaining a dynastic or elite warrior-class identity among certain groups in the Near East during the Late Bronze Age.

4.2 Geographical origins of the Indo-Iranian language

The traces of early Indo-Aryan speakers in Northern Syria positions the oldest Indo-Iranian speakers somewhere between Western Asia and the Greater Punjab, where the earliest Vedic text is thought to have been composed during the Late Bronze Age (cf. Witzel 1999: 3). In addition, a northern connection is suggested by contacts between the Indo-Iranian and the Finno-Ugric languages. Speakers of the Finno-Ugric family, whose antecedent is commonly sought in the vicinity of the Ural Mountains, followed an east-to-west trajectory through the forest zone north and directly adjacent to the steppes, producing languages across to the Baltic Sea. In the languages that split off along this trajectory, loanwords from various stages in the development of the Indo-Iranian languages can be distinguished: 1) Pre-Proto-Indo-Iranian (Proto-Finno-Ugric **kekrä* (cycle), **kesträ* (spindle), and **-teksä* (ten) are borrowed from early preforms of Sanskrit *cakrá-* (wheel, cycle), *cattra-* (spindle), and *daśa-* (10); Koivulehto 2001), 2) Proto-Indo-Iranian (Proto-Finno-Ugric **śata* (one hundred) is borrowed from a form close to Sanskrit *śatám* (one hundred), 3) Pre-Proto-Indo-Aryan (Proto-Finno-Ugric **ora* (awl), **reśmä* (rope), and **ant-* (young grass) are borrowed from preforms of Sanskrit *ā́rā-* (awl), *raśmí-* (rein), and *ándhas-* (grass); Koivulehto 2001: 250; Lubotsky 2001: 308), and 4) loanwords from later stages of Iranian (Koivulehto 2001; Korenchy 1972). The period of prehistoric language contact with Finno-Ugric thus covers the entire evolution of Pre-Proto-Indo-Iranian into Proto-Indo-Iranian, as well as the dissolution of the latter into Proto-Indo-Aryan and Proto-Iranian. As such, it situates the prehistoric location of the Indo-Iranian branch around the southern Urals (Kuz'mina 2001).

4.3 Post-steppe contacts with the Bactria-Margiana Archaeological Complex

Between the likely northern steppe homeland and the attestation of the Indo-Iranian languages in South Asia in historical times, their speakers came into contact with an unknown language probably spoken in Central Asia. Traces of this language survive in Indo-Iranian as a layer of prehistoric non-Indo-European loanwords (Parpola 2015: 81, 82; Pinault 2003, 2006; Witzel 1995: 103). This layer, which can be dated between the pre-Indo-Aryan/Finno-Ugric contacts and the appearance of Indo-Aryan words in Mitanni, includes culturally salient terms belonging to the spheres of 1) construction, cf. Proto-Indo-Iranian **ȷ́ʰarmiya-* ((permanent) building), **ištiya-* (brick), 2) land cultivation, cf. **yavīya-* (irrigation channel), **kʰā-* (dug well), and 3) local fauna, cf. **Huštra-* (Bactrian camel), **kʰara-* (donkey), **kaćyapa-* (tortoise), and 4) religion, e.g. the divinity **Indra-* (also attested in Mitanni), **atʰarvan-*

(priest), **r̥ši-* (seer), **anću-* (Soma plant) (Lubotsky 2001, 2010). Coming from the culturally and environmentally dissimilar southern Ural region, Indo-Iranian speakers were presumably unfamiliar with such phenomena and borrowed the pertaining words as they were confronted by them. Speakers of both Indo-Aryan and Tocharian, another Indo-European language spoken ca. AD 500–1000 in Northwest China, probably became acquainted with the domesticated donkey (first domesticated in Africa, cf. Parpola and Janhunen 2011; Rossel et al. 2008) through speakers of this unknown language, which served as the mediator between West Semitic *ḫāru* (donkey) (Streck 2011: 367) in Mesopotamia, and Proto-Indo-Iranian **kʰara-* (donkey) and Tocharian B *koro** (mule) (Pinault 2008: 392–393) in Central Asia.

The Bactria-Margiana Archaeological Complex (BMAC) as discussed by Sarianidi 1976 would constitute a plausible material culture analogue for the unknown language identified above (Lubotsky 2001, 2010; Witzel 2003). The linguistic makeup of BMAC and the preceding Namazga culture is unknown, but the semantics of the aforementioned non-Indo-European elements point to a language spoken by an urbanized agrarian society with a Central Asian fauna. It has been suggested on cultural and archaeological grounds that Indo-Iranian-speaking pastoral nomads prior to their spread further south interacted with the irrigation farmers of the BMAC towns (see Outram et al. 2018).

From around 1800 BCE, BMAC settlements certainly decrease sharply in size, and although BMAC-style ceramic wares continue, Andronovo pottery appears both inside urban centres and temporary pastoral campsites, which existed around BMAC sites in the hundreds (Anthony 2007: 452). This period probably marks the initial stages of agriculturalist-pastoralist interaction. Though the fortified settlements of the BMAC suggest that these contacts may not always have been peaceful (Lamberg-Karlovsky 2005: 161), agriculturalists and pastoralists would have profited from a shared mixed-subsistence economy. It has been hypothesized on the basis of palaeoethnobotanical evidence that herd animals were allowed to graze on the stubble of agricultural fields, indicating an aspect of non-hostile interaction between mobile pastoralists and settled farmers (Spengler 2014: 808, 816). In such a setting of both extensive and intensive cultural encounters, linguistic contact would be almost inevitable.

4.4 Later linguistic contacts in South Asia

It is beyond doubt that the languages of the Indo-Aryan group have been in contact with non-Indo-European languages within South Asia. However, the identification of such languages and the date of the contact are controversial.

In Indo-Aryan, a second layer of loanwords similar to those thought to originate in the BMAC is found that is absent from the Iranian languages. This layer may have been absorbed by Vedic at a later stage, i.e. after its speakers had lost direct contact with the predecessors of the Iranian languages and had begun settling in South Asia. It is therefore plausible that one of the languages spoken in the Greater Punjab prior to the arrival of Indo-Aryan speakers was similar to that spoken in the towns of Central Asia (Lubotsky 2001: 306). This would in turn point to a pre-Indo-European dispersal of a BMAC language to the Indian subcontinent.

Influence from a language of the Munda family has been posited by Kuiper and Witzel 2003. The Munda languages, spoken in central and eastern India, many clustering in Odisha and Jharkhand, form a subgroup of the larger Austro-Asiatic language family and are not genealogically related to Indo-European or Indo-Iranian. Kuiper argued that a large number of Indic words, starting from the oldest variety of the language, Rig Vedic, but continuing into later stages of Sanskrit, derives from a preform of Munda that he called Proto-Munda (1948) or Para-Munda, meaning that a language similar but not identical to Proto-Munda was the source. He also noted structural elements from Munda, such as particular sound alternations and combinations as well as prefixes and suffixes (1991). Kuiper’s theory has been accepted by Witzel (e.g. 1999: 6–10, 36–39) but has been criticized by others (e.g. Anderson 2008: 5; Osada 2006; Parpola 2015: 165).

A Dravidian influence on Sanskrit is more widely accepted (e.g. Burrow 1955: 397–398; Parpola 2015; Witzel 1999). The Dravidian languages form a family of their own and are all spoken in southern and eastern India, except Brahui, which is spoken in Pakistan. Witzel 1999: 5, who recognizes influence from both Munda and Dravidian in Rig Vedic, notes that the Munda influence begins slightly earlier than that of Dravidian (see also Zvelebil 1972).

4.5 Steppe ancestry in South Asia

The West Eurasian genetic component in South Asians can be modelled as a two-step influx from the north. The first wave, which we propose was a population genetically similar to the Early Bronze Age Namazga ancestry, introduced EHG ancestry into South Asia. The second wave also introduced EHG ancestry, but was mixed with European farmer DNA, and matches the signal traced in the Sintashta and Andronovo cultures. While the first wave cannot be linked to any known Indo-European language, the second wave coincides archaeologically with the expansion of chariotry from the southern Urals to Syria and the Indian subcontinent and linguistically with the spread of the Indo-Iranian languages. Linguistic interaction between the first and second waves can be connected to a layer of non-Indo-European vocabulary in the Indo-Iranian languages, likely reflecting contact between Namazga-derived BMAC agriculturalists and intrusive pastoralists from the northern Steppe Zone.

5. Discussion

We modify the linguistic “Steppe Hypothesis” using the new archaeological DNA presented in Damgaard et al. 2018 that traces ancestry and human mobility which we link to the dispersal of the Indo-European Anatolian, Tocharian and Indo-Iranian language families. We further test the “Steppe Hypothesis” by matching the distribution of West Eurasian ancestry in the Bronze Age against the spread of the three Indo-European branches to Anatolia, Inner Asia and South Asia.

We conclude that the EHG-related steppe ancestry found in individuals of period III Namazga culture and in modern-day populations on the Indian subcontinent cannot be linked to an Early Bronze Age intrusion of the Indo-Iranian languages in Central and South Asia associated with the Yamnaya culture. The spread of these languages may instead have been driven by movements of groups associated with the Sintashta/Andronovo culture, who were carriers of a West Eurasian genetic signature similar to the one found in individuals associated with the Corded Ware culture in Europe and who probably spread with LBA pastoral-nomads from the South Ural Mountains. Archaeologically, this wave of LBA Steppe ancestry is dated to the period after 2000 BCE when chariotry was adopted across much of Eurasia. The linguistic evidence from the reconstructed Indo-Iranian proto-language as well as the diffusion of Proto-Indo-Aryan terminology related to chariotry suggests that the speakers of Indo-Iranian took part in the proliferation of this technology to LBA Syria and Northwest India.

In Inner Asia, the previously suggested connection between the Yamnaya and Afanasievo cultures is further strengthened by the genetic ancestry of the individual coming from the intermediate site at Karagash. The Afanasievo culture is currently the best archaeological proxy for the linguistic ancestors to the speakers of the Tocharian languages.

Furthermore, our genetic data cannot confirm a scenario in which the introduction of the Anatolian Indo-European languages into Anatolia was associated with the spread of EBA Yamnaya West Eurasian ancestry. The Anatolian samples contain no discernible trace of steppe ancestry at present. The combined linguistic and genetic evidence therefore have important implications for the “Steppe Hypothesis” in Southwest Asia.

First, the lack of genetic indications for an intrusion into Anatolia refutes the classical notion of a Yamnaya-derived mass invasion or conquest. However, it does fit the recently developed consensus among linguists and historians that the speakers of the Anatolian languages established themselves in Anatolia by gradual infiltration and cultural assimilation.

Second, the attestation of Anatolian Indo-European personal names in 25th century BCE decisively falsifies the Yamnaya culture as a possible archaeological horizon for PIE-speakers prior to the Anatolian Indo-European split. The period of Proto-Anatolian linguistic unity can now be placed in the 4th millennium BCE and may have been contemporaneous with e.g. the Maykop culture (3700–3000 BCE), which influenced the formation and apparent westward migration of the Yamnaya and maintained commercial and cultural contact with the Anatolian highlands (Kristiansen et al. 2018). Our findings corroborate the Indo-Anatolian Hypothesis, which claims that Anatolian Indo-European split off from Proto-Indo-European first and that Anatolian Indo-European represents a sister rather than a daughter language. Our findings call for the identification of the speakers of Proto-Indo-Anatolian as a population earlier that the Yamnaya and late Maykop cultures.

REFERENCES

Adrados, F. R. 2007. Must we again postulate a unitary and uniform Indo-European? *Indogermanische Forschungen* 112: 1–25.

Allentoft, M. et. al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522: 167–72. DOI: 10.1038/nature14507.

Anderson, G. D. S. 2008. Introduction to the *Munda Languages*. In: G. D. S. Anderson (ed.), *The Munda languages.* London: Routledge, 1–10.

Anthony, D. W. 1995. Horse, wagon & chariot: Indo-European languages and archaeology. *Antiquity* 69: 554–65.

Anthony, D. W. 2007. *The horse, the wheel, and language: How Bronze-Age riders from the Eurasian Steppes shaped the modern world.* Princeton: Princeton University Press.

Anthony, D. W. and D. Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1: 199–219.

Archi, A. 2011. In Search of Armi. *Journal of Cuneiform Studies* 63: 5–34.

Barjamovic, G. 2011. *A Historical Geography of Anatolia in the Old Assyrian Colony Period*. Copenhagen: Museum Tusculanums Forlag.

Beekes, R. S. P. 2011. *Comparative Indo-European Linguistics. An introduction.* Second edition, revised and corrected by Michiel de Vaan. Amsterdam: John Benjamins.

Bilgiç, E. 1954. *Die Einheimischen Appellativa der kappadokischen Texte und ihre Bedeutung für die anatolischen Sprachen*. Ankara: Türk Tarih Kurumu Basımevi.

Bonechi, M. 1990. Aleppo in età arcaica; a proposito di un'opera recente. *Studi Epigrafici e Linguistici sul Vicino Oriente Antico* 7: 15–37.

Bonechi, M. 2016. Thorny Geopolitical Problems in the Palace G Archives. The Ebla Southern Horizon. Part One: The Middle Orontes Basin. In: Dominique Parayre (ed.), *Le fleuve rebelle Géographie historique du moyen Oronte d'Ebla à l'époque médiévale* (SYRIA Supp. IV). Beyrouth: Presses de l'Ifpo, 29–88.

Bouckaert et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337: 957–60.

Burrow, Th. 1955. *The Sanskrit Language*. London: Faber and Faber.

Catagnoti, A. 2012. *La grammatica della lingua di Ebla* (Quaderni di Semitistica 29). Firenze: Dipartimento di Scienze dell’Antichità.

Chang et al. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91: 194–244.

Chernykh, E. N. 1992. *Ancient metallurgy in the USSR*. Cambridge: Cambridge University Press.

Damgaard et al. 2018. The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia. *Science*.

Dassow, E. von. 2008. State and Society in the Late Bronze Age. Alalah under the Mittani Empire (Studies on the Civilization and Culture of Nuzi and the Hurrians 17). Bethesda: University Press of Maryland.

Dercksen, J. G. 2007. On Anatolian Loanwords in Akkadian Texts from Kültepe. *Zeitschrift für Assyriologie* 97: 26–46.

Del Monte, G. and J. Tischler. 1978. *Die Orts- und Gewässernamen der hethitischen Texte* (Répertoire Géographique des Textes Cunéiformes 6). Wiesbaden: Reichert.

Eidem, J. 2001. The Shemsharra Archives, Vol. 1: The Letters (Det Kongelige Danske Videnskabernes Selskab, Historisk-filosofiske Skrifter 23). Copenhagen: Det Kongelige Danske Videnskabernes Selskab.

Eidem, J. 2014. The Kingdom of Šamšī-Adad and its Legacies. In: E. Cancik-Kirschbaum, N. Brisch, and J. Eidem (eds.), *Constituent, Confederate, and Conquered Space: The Emergence of the Mittani State*. Boston/Berlin: de Gruyter, 137–46.

Fussman, G. 1972. *Atlas linguistique des parlers dardes et kafirs. II. Commentaire.* Paris: École française d’extrême orient.

Gamkrelidze, T. V. and V. V. Ivanov. 1995. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture*. 2 vols. Berlin: de Gruyter.

Gimbutas, Marija. 1965. *Bronze Age cultures in Central and Eastern Europe*. The Hague/London: Mouton.

Goedegebuure, P. 2008. Central Anatolian languages and language communities in the Colony period: A Luwian-Hattian symbiosis and the independent Hittites. In: J. G. Dercksen (ed.), *Anatolia and the Jazira during the Old Assyrian Period* (Publication de l’Institut Historique et Archéologique Néerlandais de Stamboul 111. Old Assyrian Archives

Studies 3). Leiden: Nederlands Instituut voor het Nabije Oosten, 137–80.

Hiebert, T. H. 1994. *Origins of the Bronze Age Oasis Civilization in Central Asia.* Cambridge, MA: Peabody Museum of Archaeology and Ethnology.

Hrozný, B. 1915. Die Lösung des hethitischen Problems. *Mitteilungen der Deutschen Orient-Gesellschaft* 56: 17–50.

Kloekhorst, A. The Anatolian stop system and the Indo-Hittite Hypothesis. *Indogermanische Forschungen* 121: 213–47.

Knudtzon, J. A. 1902. *Die zwei Arzawa-briefe, die ältesten urkunden in indogermanischer sprache*. Leipzig: Hinrichs.

Koivulehto, J. 2001. The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. In: Chr. Carpelan, A. Parpola, and P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations.* Helsinki: Suomalais-Ugrilainen Seura, 235–63.

Koppen, F. van. 2017. The Early Kassite Period. In A. Bartelmus and K. Sternitzke. *Karduniaš. Babylonia Under the Kassites* (Ergänzungsbände zur Zeitschrift für Assyriologie und Vorderasiatische Archäologie 11/1). Boston/Berlin: de Gruyter, 45–91.

Korenchy, É. 1972. *Iranische Lehnwörter in den obugrischen Sprachen.* Budapest: Akadémiai Kiadó.

Kortlandt, F. 1990. The spread of the Indo-Europeans. *Journal of Indo-European Studies* 18: 131–40.

Kristiansen, K. 2005. The rise of Bronze Age society: travels, transmissions and transformations. Cambridge: Cambridge University Press.

Kristiansen, K. et al. 2017. Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity* 91: 334–47.

Kristiansen et al. 2018. Archaeological supplement A to Damgaard et al. 2018: discussion of the archaeology of the Caucasus, Anatolia, Central and South Asia 4000–1500 BCE. *bioRxiv* DOI XXX.

Kuiper, F. B. J. 1948. *Proto-Munda words in Sanskrit.* Amsterdam: Noord-Hollandsche Uitgeversmaatschappij.

Kuiper, F. B. J. 1991. *Aryans in the Rigveda.* Amsterdam: Rodopi.

Kuz’mina, E. E. 2001. Contacts between Finno-Ugric and Indo-Iranian speakers in the light of archaeological, linguistic and mythological data. In: Chr. Carpelan, A. Parpola, and P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations.* Helsinki: Suomalais-Ugrilainen Seura, 289–300.

Kuz’mina, E. E. 2007. *The origin of the Indo-Iranians.* Leiden: Brill.

Kuznetsov, P. F. 2006. The emergence of Bronze Age chariots in Eastern Europe. *Antiquity* 80: 638–45.

Lamberg-Karlovsky, C. C. 2005. Language and archaeology: The case of the Bronze Age Indo-Iranians. In: Edwin Bryant and Laurie L. Patton (eds.), *The Indo-Aryan Controversy. Evidence and inference in Indian history*. Abingdon: Routledge, 142–78.

Laroche, E. 1966. *Les noms des Hittites* (Études linguistiques 4). Paris: C. Klincksieck.

Laroche, E. 1981. Les noms des Hittites: supplement. *Hethitica* 4: 3–58.

Larsen, M. T. 2015. *Ancient Kanesh*. *A Commercial Colony in Bronze Age Anatolia*. Cambridge: Cambridge University Press.

Larsen, M. T. and Lassen, A. W. 2014. Cultural Exchange at Kültepe. In: M. Kozuh (ed.), *Extraction and Control*. *Studies in Honor of Matthew W. Stolper* (Studies in Ancient Oriental Civilization 68). Chicago: University of Chicago, 170–88.

Lehrman, A. 1998. *Indo-Hittite Redux*. Moscow: Paleograph.

Lubotsky, A. M. 2001. The Indo-Iranian substratum. In: Chr. Carpelan, A. Parpola, and P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations.* Helsinki: Suomalais-Ugrilainen Seura, 301–17.

Lubotsky, A. M. 2010. Kto byli žiteli Gonura i na kakom jazyke oni govorili? In: P. M. Kozhin, M. F. Kosarev, and N. A. Dubova (eds.), *On the track of uncovering a civilization. A volume in honor of the 80th anniversary of Victor Sarianidi.* Sankt-Peterburg: Aletheia, 18–22.

Malandra, W. W. 1991. Chariot. In: *Encyclopædia Iranica* V: 377–80. [http://www.iranicaonline.org/articles/chariot-av] (Last Updated: October 13, 2011).

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

Mallory, J. P. 1998. A European perspective on Indo-Europeans in Asia. In: V. H. Mair (ed.), *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia, I: Archeology, Migration and Nomadism, Linguistics.* Washington D.C.: The Institute for the Study of Man, 175–201.

Mallory, J. P. 2015. *The problem of Tocharian origins: An archaeological perspective.* (Sino-Platonic Papers 259) Philadelphia: University of Pennsylvania. [www.sino-platonic.org]

Mallory, J. P. and D. Q. Adams. 1997. *Encyclopedia of Indo-European Culture*. London: Fitzroy Dearborn Publishers.

Mallory, J. P. and V. H. Mair. 2000. *The Tarim mummies. Ancient China and the mystery of the earliest peoples from the west.* London: Thames and Hudson.

Mayrhofer, M. 1966. *Die Indo-Arier im alten Vorderasien. Mit einer analytischen Bibliographie.* Wiesbaden: Harrassowitz.

Mayrhofer, M. 1974. *Die Arier im vorderen Orient—ein Mythos?* (Österreichische Akademie der Wissenschaften. Philosophisch-Historische Klasse 294/3). Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Mayrhofer, M. 1982. Welches Material aus dem Indo-Arischen von Mitanni verbleibt für eine selektive Darstellung? In: E. Neu (ed.), *Investigationes philologicae et comparativae: Gedenkschrift für Heinz Kronasser*. Wiesbaden: Harrassowitz, 72–90.

Melchert, H. C. 2003. Prehistory. In: H.C. Melchert (ed.), *The Luwians*. Leiden: Brill, 8–26.

Melchert, H. C. 2011. Indo-Europeans. In: Sharon Steadman and Gregory McMahon (eds.), *The Oxford Handbook of Ancient Anatolia. 10,000—323 B.C.E.* Oxford/New York: Oxford University Press, 704–16.

Melchert, H. C. 2017. Anatolian. In: M. Kapović, A. G. Ramat, and P. Ramat (eds.), *The Indo-European Languages*. Second edition. New York: Routledge, 171–201.

Melchert, H. C. (fthc). The Position of Anatolian. In: A. Garrett and M. Weiss (eds.), *Handbook of Indo-European Studies*. Oxford: Oxford University Press.

Morgenstierne, G. 1973. Die Stellung der Kafirsprachen. In: *Irano-Dardica* (Beiträge zur Iranistik 5). Wiesbaden: Reichert, 327–43.

Oettinger, N. 2002. Indogermanische Sprachträger lebten schon im 3. Jahrtausend v.Chr. in Kleinasien. Die Ausbildung der anatolischen Sprachen. In: Helga Willinghöfer (ed.), *Die Hethiter und ihr Reich. Das Volk der tausend Götter*. Stuttgart/Bonn: Theiss-Verlag, 50–55.

Oettinger, N. 2013–2014. Die Indo-Hittite-Hypothese aus heutiger Sicht. *Münchener Studien zur Sprachwissenschaft* 67(2): 149–76.

Osada, T. 2006. How many Proto-Munda words in Sanskrit? With special reference to agricultural vocabulary. In: T. Osada (ed.), *Proceedings of the Pre-Symposium of RIHN and the 7th ESCA Harvard-Kyoto Roundtable.* Kyoto: Research Institute for Humanity and Nature, 151–74.

Outram et al. 2018. Archaeological supplement B to Damgaard et al. 2018: discussion of the archaeology of Central Asian and East Asian Neolithic to Bronze Age hunter-gatherers and early pastoralists, including consideration of horse domestication. *bioRxiv* DOI XXX.

Parpola, A. 1995. The problem of the Aryans and the Soma: the archaeological evidence. In: George Edosy (ed.), *The Indo-Aryans of ancient South Asia language: language material culture and ethnicity*. Berlin/New York: de Gruyter, 353–81.

Parpola, A. 2004-2005. The Nāsatyas, the Chariot and Proto-Aryan Religion. *Journal of Indological Studies* 16/17: 1–63.

Parpola, A. 2015. *The roots of Hinduism*. Oxford: Oxford University.

Parpola, A. and J. Parpola. 2011. On the Asiatic wild asses and their vernacular names. In: T. Osada and H. Endo (eds.), *Linguistics, archaeology and the human past: Occasional paper 12*. Kyoto: Research Institute for Humanity and Nature, 59–124.

Peyrot, M. 2017. Tocharian: An Indo-European language from China. In: J. M. Kelder, S. P. L. de Jong and A. Mouret (eds.), *Aspects of globalisation.*

Mobility, exchange and the development of multi-cultural states. Leiden: LURIS, 12–17. [http://hdl.handle.net/1887/48072].

Pinault, G.-J. 2003. Sanskrit *kalyāṇa-* interpreté à la lumière des contacts en Asie Centrale. *Bulletin de la Société de Linguistique* 98: 123–61.

Pinault, G.-J. 2006. Further links between the Indo-Iranian substratum and the BMAC language. In: Bertil Tikkanen and Heinrich Hettrich (eds.), *Themes and Tasks in Old and Middle Indo-Aryan Linguistics.* Delhi: Motilal Banarsidass, 167–96.

Pinault, G.-J. 2008. *Chrestomathie tokharienne. Textes et Grammaire*. Leuven: Peeters.

Puhvel, J. 1994. Anatolian: Autochthon or Interloper? *Journal of Indo-European Studies* 18: 251–63.

Renfrew, A. C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Pimlico.

Renfrew, A. C. 1999. Time depth, convergence theory, and innovation in Proto-Indo-European: “Old Europe” as a PIE linguistic area. *Journal of Indo-European Studies* 27(3–4): 257–93.

Rieken, E. 2009. Der Archaismus des Hethitischen: Eine Bestandsaufnahme. *Incontri Linguistici* 32: 37–52.

Rossel, S., F. Marshall, J. Peters, T. Pilgram, M. D. Adams and D. O’Connor. 2012. Domestication of the donkey: Timing, processes, and indicators. PNAS 105/10, 3715-3720.

Sarianidi, V. I. 1976. *Issledovanija pamjatnikov Dashlyiskogo Oazisa* (Drevnii Baktria 1). Moscow: Akademia Nauk.

Schwemer, D. 2005–2006. Lehnbeziehungen zwischen dem Hethitischen und dem Akkadischen. *Archiv für Orientforschung* 51: 220–34.

Simons, G. F. and Ch. D. Fennig (eds.). 2017. *Ethnologue: Languages of the World*. Twentieth edition. Dallas: SIL International.

Spengler, R. N. et al. 2014. Agriculturalists and pastoralists: Bronze Age economy. *Vegetation History and Archaeobotany* 23: 805–20. DOI: 10.1007/s00334-014-0448-0.

Stefanini, R. 2002. Toward a diachronic reconstruction of the linguistic map of Ancient Anatolia. In: S. de Martino and F. Pecchioli-Daddi (eds.), *Anatolia Antica: Studi in memoria di Fiorella Imparati* (Eothen 11). Firenze: LoGisma, 783–806.

Steiner, G. 1981. The Role of the Hittites in Ancient Anatolia. *Journal of Indo-European Studies* 9: 150–73.

Steiner, G. 1990. The immigration of the first Indo-Europeans into Anatolia reconsidered. *Journal of Indo-European Studies* 18: 185–214.

Strand, R. F. 1973. Notes on the Nûristânî and Dardic “Languages. *Journal of the American Oriental Society* 93: 297–305.

Teufer, M. 2012. Der Streitwagen: eine “indo-iranische” Erfindung? Zum Problem der Verbindung von Sprachwissenschaft und Archäologie. *Archäologische Mitteilungen aus Iran und Turan* 44: 271–312.

Thieme, P. 1960. The ‘Aryan Gods’ of the Mitanni Treaties. *Journal of the American Oriental Society* 80: 301–17.

Vadeckaja, È. B. 1986. *Arxeologičeskie pamjatniki v stepjax srednego Jeniseja*. Leningrad: Nauka.

Vadeckaja, È. B., A. V. Poljakov and N. F. Stepanova. 2014. *Svod pamjatnikov afanas’evskoj kul’tury.* Barnaul: Azbuka.

Wáng Bǐnghuá. 1999. Luóbù nào'ěr gǔ shī — The mummies of Lopnur. In: Wáng Bǐnghuá (ed.), *Xīnjiāng gǔshi: gǔdài Xīnjiāng jūmín jí qí wénhuà — The ancient corpses of Xinjiang: the peoples of ancient Xinjiang and their culture.* Ürümči: Xīnjiāng Rénmín Chūbǎnshè.

Watkins, C. 2001. An Indo-European linguistic area and its characteristics: Ancient Anatolia. Areal diffusion as a challenge to the comparative method? In: A. Y. Aikhenvald and R. M. W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press, 44–63.

Weeden, M. 2011. *Hittite Logograms and Hittite Scholarship* (Studien zu den Boğazköy-Texten 54). Wiesbaden: Harrassowitz.

Wilhelm, G. 2008. Hurrians in Kültepe texts. In: J. G. Dercksen (ed.), *Anatolia and the Jazira during the Old Assyrian Period* (Publication de l’Institut Historique et Archéologique Néerlandais de Stamboul 111. Old Assyrian Archives Studies 3). Leiden: Nederlands Instituut voor het Nabije Oosten, 181–94.

Winn, S. M. M. 1981. Burial evidence and the kurgan culture in Eastern Anatolia c. 3000 BC: An interpretation. *Journal of Indo-European Studies* 9: 113–18.

Winters, R. in prep. “Negotiating Exchange. Trade, Travel and Diplomacy in the Ebla Archives.” Unpublished PhD-dissertation, Harvard University.

Witzel, M. 1995. Early Indian history: linguistic and textual parametres. In: G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia. Language, Material culture and ethnicity.* Berlin/New York: de Gruyter, 192–205.

Witzel, M. 1999. Substrate Languages in Old Indo-Aryan (Ṛgvedic, Middle and Late Vedic). *Electronic Journal of Vedic Studies* 5(1): 1–68. doi: 10.11588/ejvs.1999.1.828.

Witzel, M. 2001. Autochthonous Aryans? The Evidence from Old Indian and Iranian Texts. *Electronic Journal of Vedic Studies* 7(3): 1–118. doi: 10.11588/ejvs.2001.3.830.

Witzel, M. 2003. Linguistic evidence for cultural exchange in prehistoric Western Central Asia. *Sino-Platonic Papers* 129: 1–70.

Witzel, M. 2005. Indocentrism. In: E. Bryant and L. L. Patton (eds.), *The Indo-Aryan Controversy. Evidence and inference in Indian history*. Abingdon: Routledge, 341–404.

Zehnder, T. 2010. *Die hethitischen Frauennamen*. Wiesbaden: Harrassowitz.

Zvelebil, K. 1972. The descent of the Dravidians. *International Journal of Dravidian Linguistics* 1(2): 56–63.

Introduction

According to the commonly accepted "Steppe Hypothesis," the initial spread of Indo-European (IE) languages into both Europe and Asia took place with migrations of Early Bronze Age Yamnaya pastoralists from the Pontic–Caspian steppe. This is believed to have been enabled by horse domestication, which revolutionized transport and warfare. While in Europe there is much support for the Steppe Hypothesis, the impact of Western steppe pastoralists in Asia, including Anatolia, remains less well understood, with limited archaeological evidence for their presence. Furthermore, the earliest secure evidence of horse husbandry comes from the Botai culture of Central Asia, while direct evidence for Yamnaya equestrianism remains elusive.

Rationale

We investigate the genetic impact of Early Bronze Age migrations into Asia and interpret our findings in relation to the Steppe Hypothesis and early spread of IE languages. We generated whole-genome shotgun sequence data (~1-25 X average coverage) for 74 ancient individuals from Inner Asia and Anatolia as well as 41 high-coverage present-day genomes from 17 Central Asian ethnicities.

Results

We show that the population at Botai associated with the earliest evidence for horse husbandry derived from an ancient hunter-gatherer ancestry previously seen in the Upper Paleolithic Mal'ta (MA1), and was deeply diverged from the Western steppe pastoralists. They form part of a previously undescribed west-to-east cline of Holocene prehistoric steppe genetic ancestry in which Botai, Central Asians, and Baikal groups can be modeled with different amounts of Eastern hunter-gatherer (EHG) and Ancient East Asian (AEA) genetic ancestry represented by Baikal_EN.

In Anatolia, Bronze Age samples, including from Hittite speaking settlements associated with the first written evidence of IE languages, show genetic continuity with preceding Anatolian Copper Age (CA) samples and have substantial Caucasian hunter-gatherer (CHG)-related ancestry but no evidence of direct steppe admixture.

In South Asia, we identify at least two distinct waves of admixture from the west: the first occurring from a source related to the Copper Age Namazga farming culture from the southern edge of the steppe, the second by Late Bronze Age steppe groups into the northwest of the subcontinent.

Conclusions

Our findings reveal that the early spread of Yamnaya Bronze Age pastoralists had limited genetic impact in Anatolia as well as Central and South Asia. As such, the Asian story of Early Bronze Age expansions differs from that of Europe. Intriguingly, we find that direct descendants of Upper Paleolithic hunter-gatherers of Central Asia, now extinct as a separate lineage, survived well into the Bronze Age. These groups likely engaged in early horse domestication as a prey-route transition from hunting to herding, as otherwise seen for reindeer. Our findings further suggest that West Eurasian ancestry entered South Asia before and after, rather than during, the initial expansion of western steppe pastoralists, with the later event consistent with a Late Bronze Age entry of IE languages into South Asia. Finally, the lack of steppe ancestry in samples from Anatolia indicates that the spread of IE languages into that region was not associated with a steppe migration.

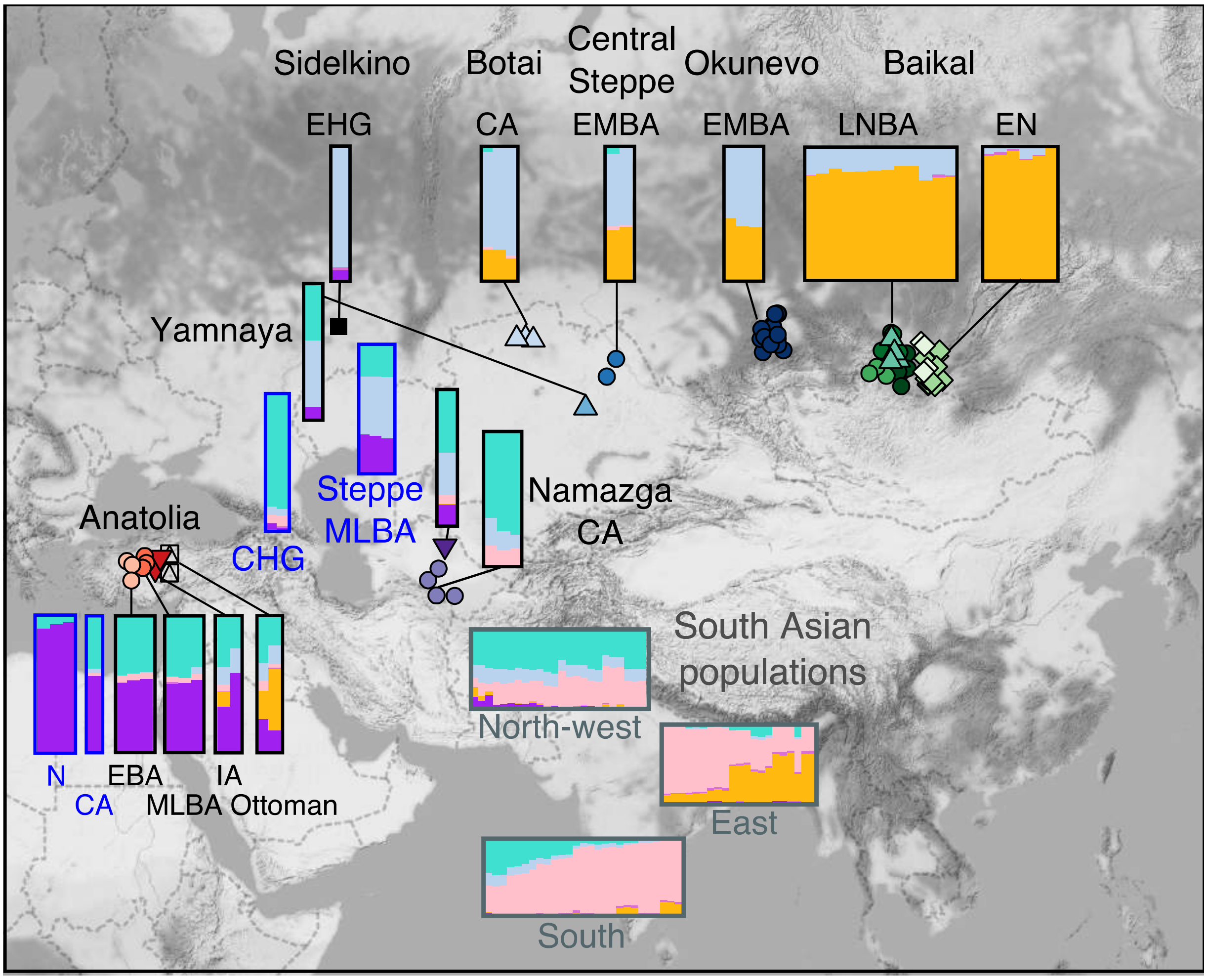
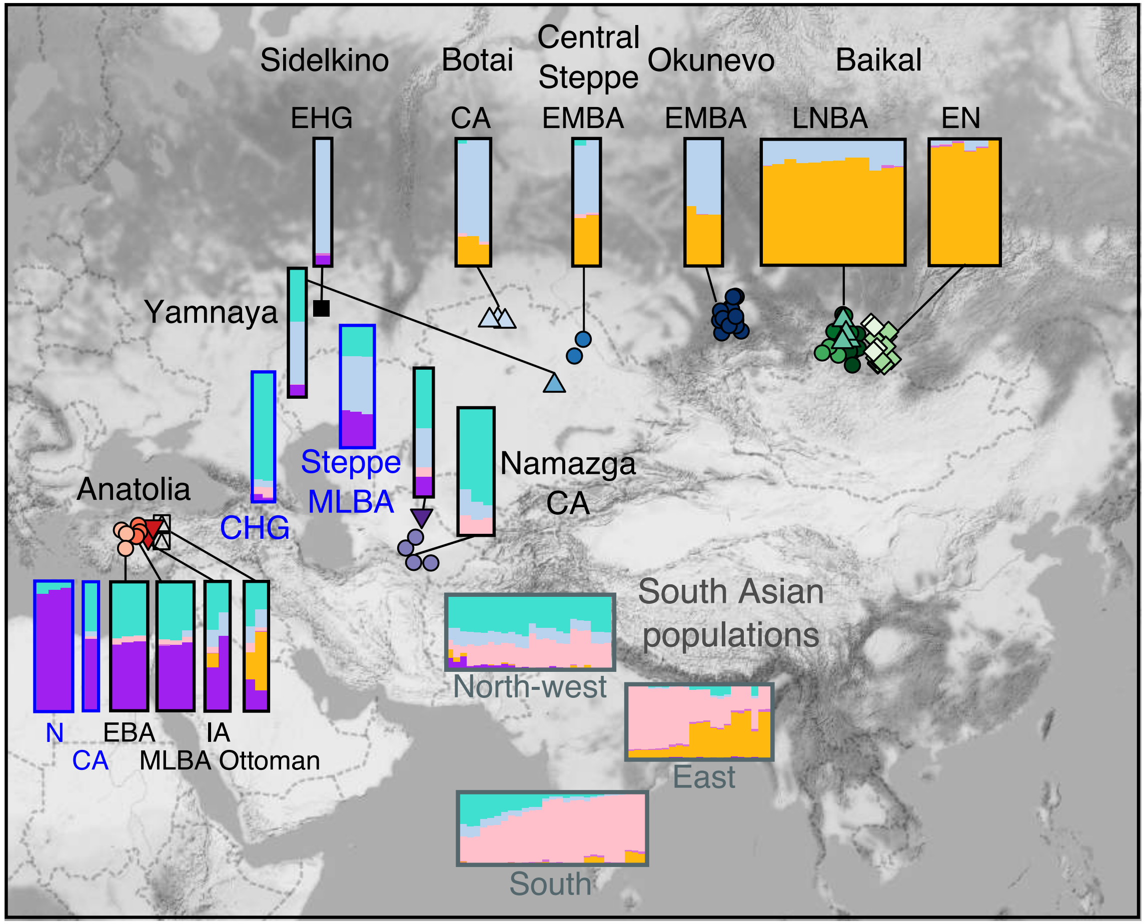


Figure Caption: Model-based admixture proportions for selected ancient and present-day individuals, assuming k=6, shown with their corresponding geographical locations. Ancient groups are represented by larger admixture plots with those sequenced in the present work surrounded by black borders, and others used for providing context with blue borders. Present-day South Asian groups are represented by smaller admixture plots with dark grey borders.


Sidelkino
Botai
Central Steppe
Okunevo
Baikal
EHG
CA
EMBA
EMBA
LNBA
EN
Yamnaya
Steppe MLBA
Namazga CA
Anatolia
CHG
South Asian populations
North-west
East
South
N
CA
EBA
MLBA
IA
Ottoman