

## RESEARCH ARTICLE SUMMARY

## PALEOGENOMICS

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

Peter de Barros Damgaard *et al.*\*

**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. Although in Europe there is much support for the steppe hypothesis, the impact of Early Bronze Age Western steppe pastoralists in Asia, including Anatolia and South Asia, 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, whereas direct evidence for Yamnaya equestrianism remains elusive.

**RATIONALE:** We investigated 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 to 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 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 identified at least two distinct waves of admixture from the west,

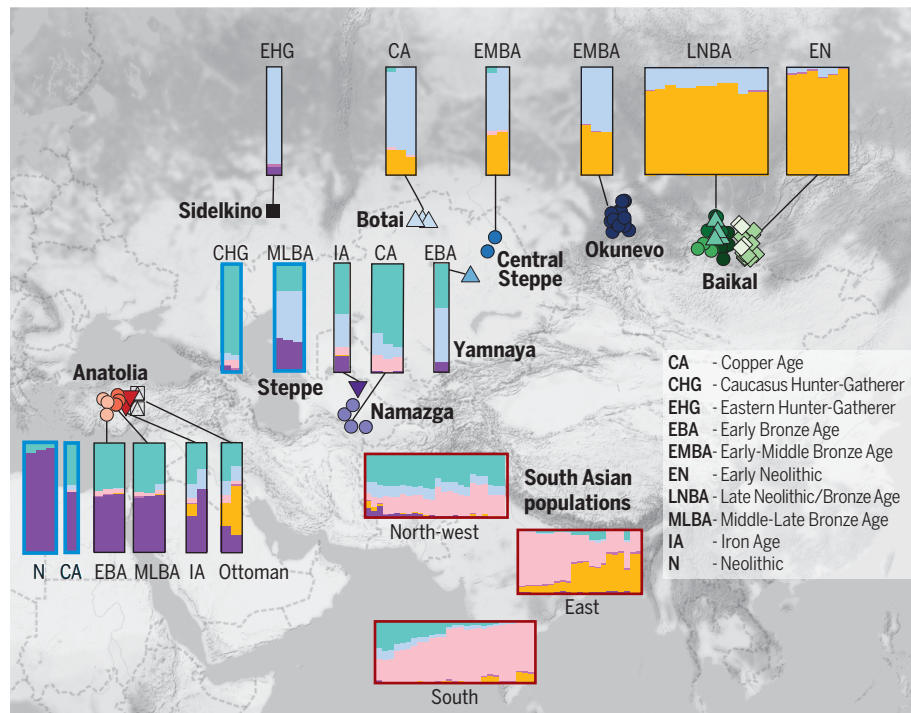
## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aar7711>

the first occurring from a source related to the Copper Age Namazga farming culture from the southern edge of the steppe, who exhibit both the Iranian and the EHG components

found in many contemporary Pakistani and Indian groups from across the subcontinent. The second came from Late Bronze Age steppe sources, with a genetic impact that is more localized in the north and west.

**CONCLUSION:** 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 pre-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 the earliest branch of IE languages into that region was not associated with a major population migration from the steppe. ■



**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 red borders.

The list of author affiliations is available in the full article online.  
\*These authors contributed equally to this work.

†Corresponding author. Email: [rd109@cam.ac.uk](mailto:rd109@cam.ac.uk) (R.D.); [ewillerslev@snm.ku.dk](mailto:ewillerslev@snm.ku.dk) (E.W.)

Cite this article as P. de Barros Damgaard *et al.*, *Science* 360, eaar7711 (2018). DOI: 10.1126/science.aar7711

## RESEARCH ARTICLE

## PALEOGENOMICS

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

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

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 analyzed 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.

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 homogeneous 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 to 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 to 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);

<sup>1</sup>Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Copenhagen, Denmark. <sup>2</sup>Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK.

<sup>3</sup>Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. <sup>4</sup>Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark.

<sup>5</sup>Leiden University Centre for Linguistics, Leiden University, Leiden, Netherlands. <sup>6</sup>Department of Near Eastern Languages and Civilizations, Harvard University, Cambridge, MA, USA. <sup>7</sup>Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, Denmark. <sup>8</sup>Shejire DNA project, Abai ave. 150/230, 050046 Almaty, Kazakhstan. <sup>9</sup>Institute of Archaeology and Steppe Civilization, Al-Farabi Kazakh National University, Almaty, 050040, Kazakhstan. <sup>10</sup>S. Toraighyrov Pavlodar State University, Joint Research Center for Archeological Studies named after A.Kh. Margulan, Pavlodar, Kazakhstan. <sup>11</sup>Department of Human Genetics, University of Chicago, Chicago, IL, USA. <sup>12</sup>Saryarkinsky Institute of Archaeology, Buketov Karaganda State University, Karaganda, 100074, Kazakhstan. <sup>13</sup>Department of Anthropology, University of Alaska, Fairbanks, AK, USA. <sup>14</sup>The Institute of Forensic Sciences, Istanbul University, Istanbul, Turkey. <sup>15</sup>Department of Genetics, Hazara University, Garden Campus, Mansehra, Pakistan. <sup>16</sup>Department of Historical Studies, University of Gothenburg, 40530 Göteborg, Sweden. <sup>17</sup>Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, St. Petersburg, Russia. <sup>18</sup>Institute for the History of Material Culture, Russian Academy of Sciences, St. Petersburg, Russia. <sup>19</sup>Japanese Institute of Anatolian Archaeology, Kaman, Kirsehir, Turkey. <sup>20</sup>Department of Archaeology, Faculty of Arts, Gazi University, Ankara, Turkey. <sup>21</sup>Center of Omic Sciences, Islamia College, Peshawar, Pakistan. <sup>22</sup>Department of Archaeology, University of Exeter, Exeter, EX4 4QE, UK. <sup>23</sup>Department of Archaeology, Hazara University, Garden Campus, Mansehra, Pakistan. <sup>24</sup>Directorate of Archaeology and Museums Government of Khyber Pakhtunkhwa, Pakistan. <sup>25</sup>Archaeological Museum Harappa at Archaeology Department Govt. of Punjab, Pakistan. <sup>26</sup>Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, Academician Lavrent'ev Ave. 17, Novosibirsk, 630090, Russia. <sup>27</sup>Department of History, Irkutsk State University, Karl Marx Street 1, Irkutsk 664003, Russia. <sup>28</sup>Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. <sup>29</sup>Department of Anthropology, University of Alberta, Edmonton, Alberta, T6G 2H4, Canada. <sup>30</sup>Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse, Université Paul Sabatier, 31000 Toulouse, France. <sup>31</sup>Departments of Integrative Biology and Statistics, University of Berkeley, Berkeley, CA, USA. <sup>32</sup>Department of Zoology, University of Cambridge, Cambridge, UK.

\*These authors contributed equally to this work.

†Corresponding author. Email: rd109@cam.ac.uk (R.D.); ewillerslev@snm.ku.dk (E.W.)



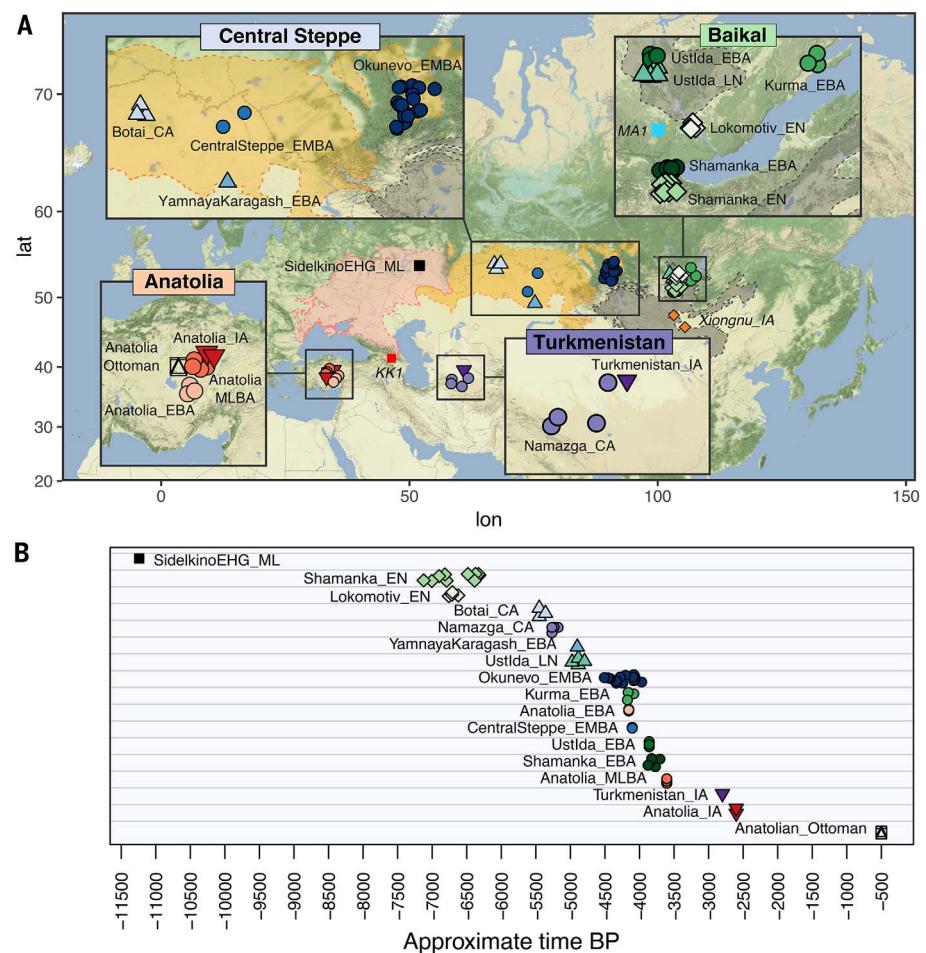
19 individuals of the Bronze Age (~2500 to 2000 BCE) Okunevo culture of the Minusinsk Basin in the Altai region (Okunevo\_EMBA; ~1X average coverage; 0.1 to 4.6X); 31 Baikal hunter-gatherer genomes (~1X average coverage; 0.2 to 4.5X) from the cis-Baikal region bordering on Mongolia and ranging in time from the Early Neolithic (~5200 to 4200 BCE; Baikal\_EN) to the Early Bronze Age (~2200 to 1800 BCE; Baikal\_EBA); 4 Copper Age individuals (~3300 to 3200 BCE; Namazga\_CA; ~1X average coverage; 0.1 to 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 to 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), and collected and genotyped 140 individuals from five 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 and table S5). Our high-coverage Yamnaya genome from Karagash is consistent with previously published Yamnaya and Afanasievo genomes, and our SidelkinoEHG\_ML 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  $D(\text{Test}, \text{Mbuti}; \text{SidelkinoEHG\_ML}, \text{EHG})$  (fig. S7) or of the form  $D(\text{Test}, \text{Mbuti}; \text{YamnayaKaragash\_EBA}, \text{Yamnaya})$  (fig. S8; additional D statistics shown in figs. S9 to S12).

### Genetic origins of local Inner Asian populations

In the Early Bronze Age, ~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 (7) and are often identified as the ancestors of the IE-speaking Tocharians of first-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 pop-

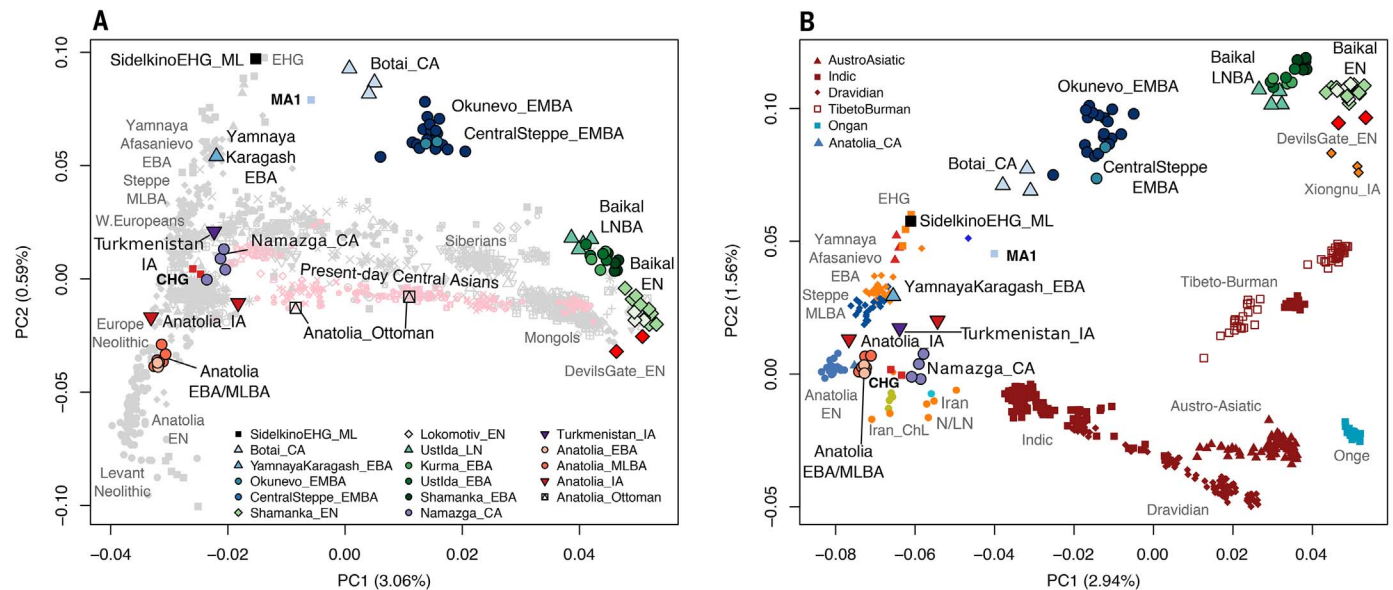


**Fig. 1. Geographic location and dates of ancient samples.** (A) Location of the 74 samples from the steppe, Lake Baikal region, Turkmenistan, and Anatolia analyzed in the present study. MA1, KK1, and Xiongnu\_IA were previously published. Geographical background colors indicate the western steppe (pink), 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.

ulations 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 to 1800 BCE) and Neolithic (~5200 to 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 and 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:  $D(\text{Baikal\_EN}, \text{Mbuti}; \text{Botai\_CA}, \text{Okunevo\_EMBA}) = -0.025$   $Z = -12$ ;  $D(\text{Baikal\_EN}, \text{Mbuti}; \text{Botai\_CA}, \text{Sholpan}) = -0.028$   $Z = -8.34$ ;  $D(\text{Baikal\_EN}, \text{Mbuti}; \text{Botai\_CA}, \text{Gregorievka}) = -0.026$   $Z = -7.1$ . 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 24,000-year-old Mal'ta (MA1), the 17,000-year-old AG-2 (26), and the ~14,700-year-old AG-3 (27) individuals from Siberia.

To investigate ancestral relationships between these populations, we used coalescent modeling with the momi (Moran Models for Inference) program (28) (Fig. 4, figs. S16 to S22, and 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



**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, whereas 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 toward 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.

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 modeled 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 ~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 ~15,000 years ago 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 ~15,000 years ago, 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 40,000 years ago, relatively soon after the peopling of Eurasia by modern humans.

Because the ANE MA1 sample comes from the same cis-Baikar region as the AEA-derived Neolithic samples analyzed here, we 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 ~5 to 11% (qpAdm) (table S12) (2), using MA1 as a source of ANE, Baikal\_EN as a source of AEA, and a set of six 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, and 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 and table S15), which could either indicate male-mediated admixture or reflect bottlenecks in the male population.

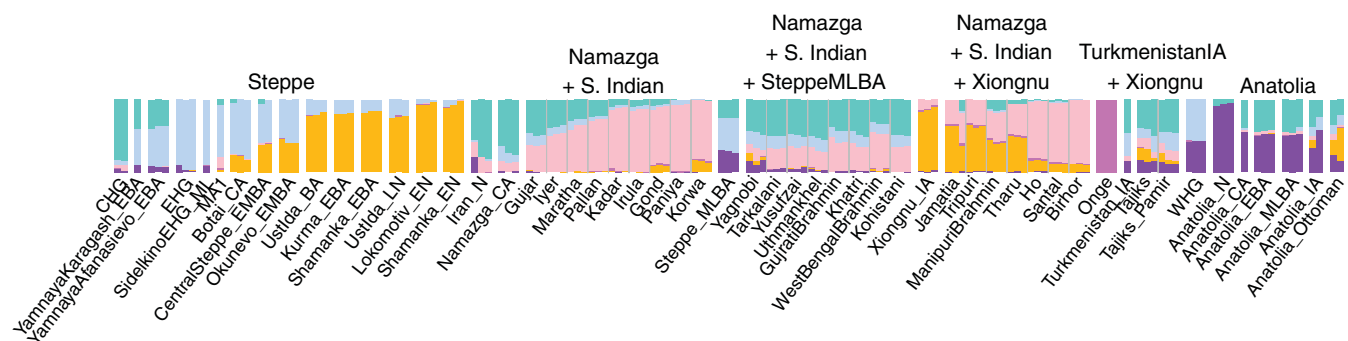
The deep population structure among the local populations in Inner Asia around the Copper Age/Bronze Age transition is in line with dis-

tinct 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 ANE 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 ANE by Early Neolithic communities of Ancient East Asian ancestry, and second, an admixture event between the latter and additional members of the ANE 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 ~3500 to 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). Although horse husbandry has been demonstrated at Botai, it is





**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.

also now clear from genetic studies that this was not the source of modern domestic horse stock (32). Some have suggested that the Botai were local hunter-gatherers who learned horse husbandry from an early eastward spread of western pastoralists, such as the Copper Age herders buried at Khvalynsk (~5150 to 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 modeled 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 ~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 sample frequency spectrum (SFS) was not significantly different from 0 ( $P = 0.18$  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  $Z = 0.45$  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  $f_4$  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). Although 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, whereas 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 and 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 toward 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 to 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 re-

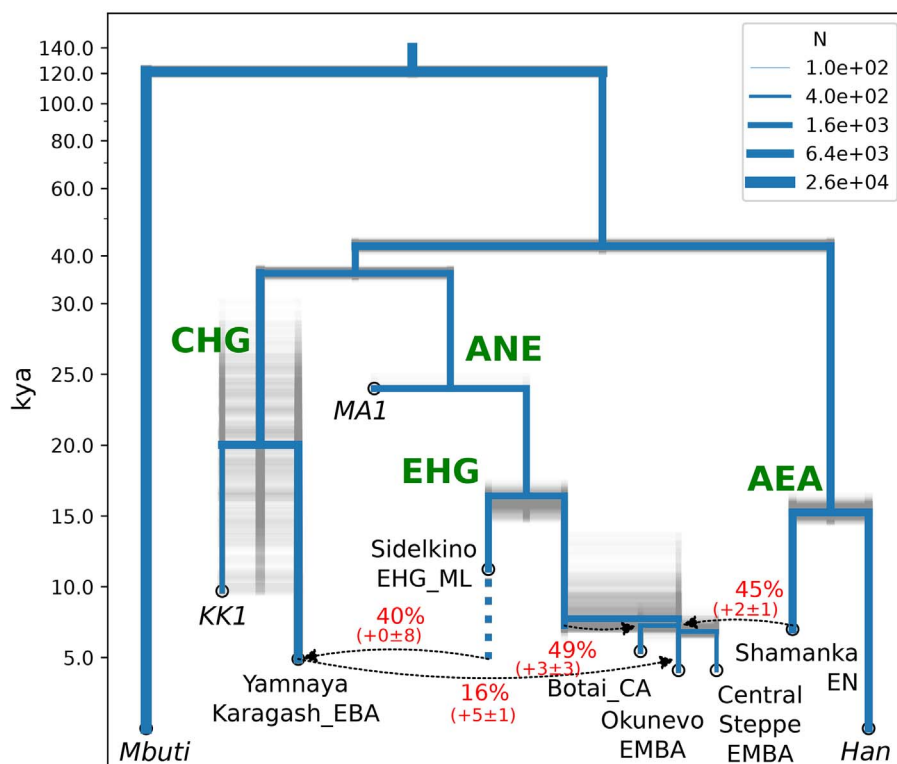
gion Bronze Age (~2200 to 1800 BCE) samples (fig. S14).

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). Notably, 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 to 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 to 1200 BCE) rather than Early Bronze Age (~3000 to 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 to 2500 BCE) steppe pastoralists Yamnaya and Afanasievo and Late



**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. The uncertainty may be underestimated here, due to simplifications or additional uncertainty in the model specification.

Bronze Age (~2300 to 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 four 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 to 3950 BCE) and substantially to Early Bronze Age (~3000 to 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,

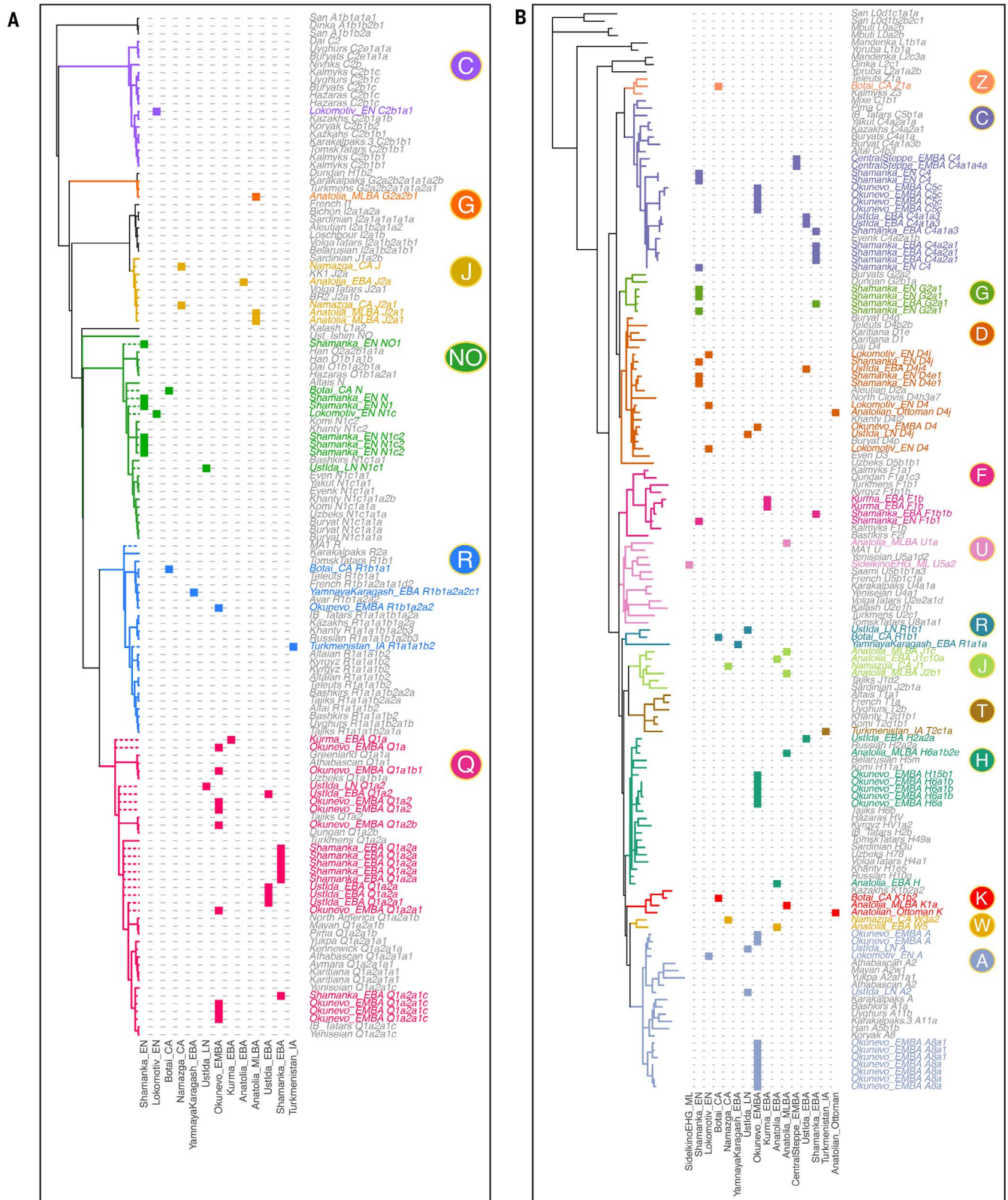
Iran\_N) does not deviate significantly from 0 ( $Z = 1.65$ ), in agreement with ADMIXTURE results (Fig. 3 and fig. S14). Moreover, a three-population qpAdm model using Iran Neolithic, EHG, and CHG as sources yields a negative admixture coefficient for CHG. This suggests that while we cannot totally reject a minor presence of CHG ancestry, steppe-related admixture most likely arrived in the Namazga population before 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 to 200 BCE) individual from the same region as Namazga (sample DA382, labeled 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 to 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 to 1200 BCE) rather than Early Bronze Age (~3000 to 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. Because 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 nine modern southern and predominantly Dravidian-speaking populations (Fig. 6, fig. S36, and tables S16 and S17). In contrast, for seven 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 to 1200 BCE) steppe source. Last, for seven southeastern Asian populations, six of which were Tibeto-Burman or Austro-Asiatic speakers, the three-population model with Late Bronze Age (~2300 to 1200 BCE) steppe ancestry was rejected, but not a model in which Late Bronze Age (~2300 to 1200 BCE) steppe ancestry was replaced with an East Asian ancestry source, as represented by the Late Iron Age (~200 BCE to 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 to 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 are 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 to 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).

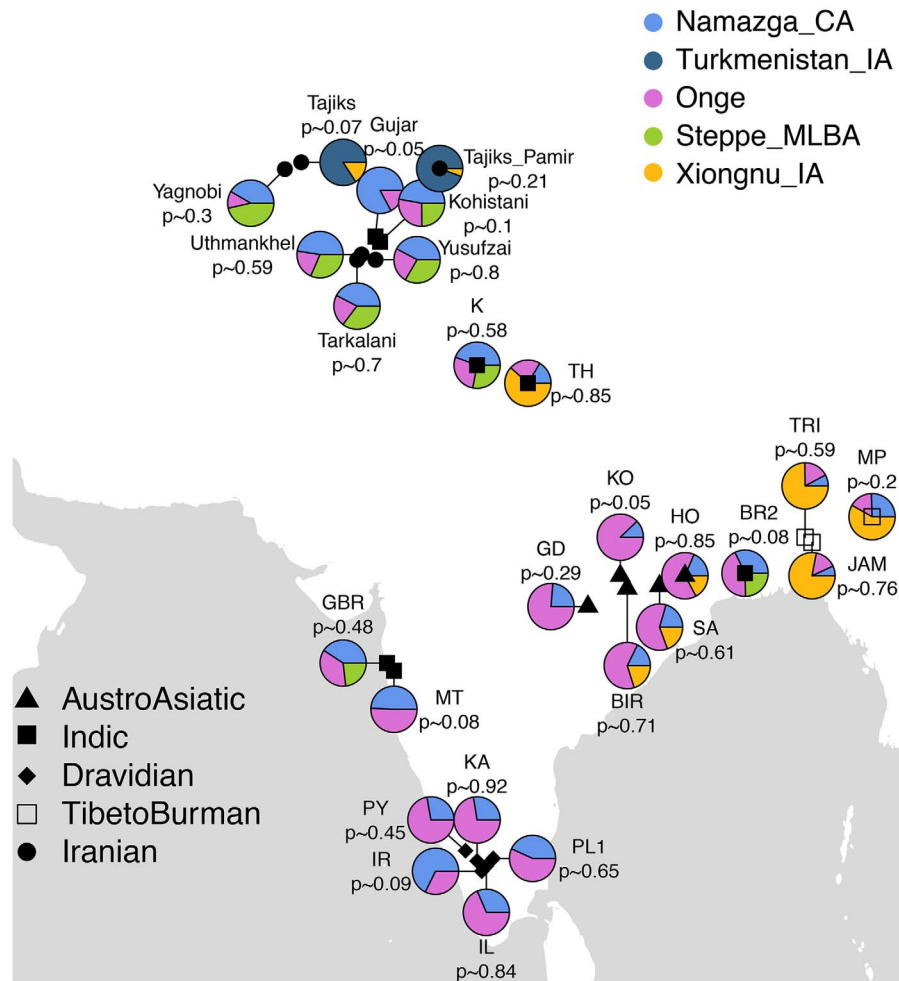
Thus, we find that ancestries deriving from four major separate sources fully reconcile the population history of present-day South Asians (Figs. 3 and 6), one anciently South Asian, one from Namazga or a related population, a third from Late Bronze Age (~2300 to 1200 BCE) steppe pastoralists, and one 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 to 1200 BCE) steppe rather than Early Bronze Age (~3000 to 2500 BCE) Yamnaya and Afanasievo admixture into South Asia. The proposal that the IE steppe ancestry arrived in the Late Bronze Age (~2300 to 1200 BCE) is also more consistent with archaeological and linguistic chronology (44, 45, 48, 49). Thus, it seems that the Yamnaya- and Afanasievo-related



**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 mitochondrial DNA (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 + Ouge, 2. Namazga\_CA + Ouge + Late Bronze Age Steppe, 3. Namazga\_CA + Ouge + Xiongnu\_IA (East Asian proxy), and 4. Turkmenistan\_IA + Xiongnu\_IA. Xiongnu\_IA were used here to represent East Asian ancestry. We observe that although South Asian Dravidian speakers can be modeled as a mixture of Ouge 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.

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 (51, 52, 53). 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 to 3000 BCE) involving movement of groups through the Balkans into Northwest Anatolia (4, 54, 55) or a Caucasian route that links language dispersal to intensified north-south population contacts facilitated by the trans-Caucasian Maykop culture ~3700 to 3000 BCE (50, 54).

Ancient DNA findings suggest extensive population contact between the Caucasus and the steppe during the Copper Age (~5000 to 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, 56). This scenario is consistent with both the Copper Age steppe (4) and the Caucasian models for the origin of the Proto-Anatolian language (57).

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 to 3700 BCE) individual from North-western 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  $D(\text{CHG}, \text{Mbuti}; \text{Anatolia\_EBA}, \text{Anatolia\_N})$  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 ( $P = 0.5$ ). This signal is not driven by Neolithic Iranian ancestry, because the result of a similar test of the form  $D(\text{Iran\_N}, \text{Mbuti}; \text{Anatolia\_EBA}, \text{Anatolia\_N})$  does not deviate from zero ( $Z = 1.02$ ). Taken together with recent findings of CHG ancestry on Crete (58), 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 (59).

Importantly, a test of the form  $D(\text{EHG}, \text{Mbuti}; \text{Anatolia\_EBA}, \text{Anatolia\_MLBA})$  supports that the Central Anatolian gene pools, including those sampled from settlements thought to have been inhabited by Hittite speakers, were not affected 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 fifth millennium BCE that seem to come to an end in the first half of the fourth millennium BCE with the village-based egalitarian Kura-Araxes' society (60, 61), 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 (62). Additionally, and in agreement with the later historical record of the region (63), 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 (64).

### Discussion

For Europe, ancient genomics have revealed extensive population migrations, replacements,



and admixtures from the Upper Paleolithic to the Bronze Age (1, 2, 27, 65, 66), with a strong influence across the continent from the Early Bronze Age (~3000 to 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 to 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 affected 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. Work on ancient horse genomes (32) indicates that Botai horses were not the main source of modern domesticates, which 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 before 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 to 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 ~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 onward (49, 67).

In Anatolia, our samples do not genetically distinguish Hittite and other Bronze Age Anatolians from an earlier Copper Age sample (~3943 to 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 (68) 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 [unlike the findings in (4)]. Our findings are thus consistent

with historical models of cultural hybridity and “middle ground” in a multicultural and multilingual but genetically homogeneous Bronze Age Anatolia (69, 70).

Current linguistic estimations converge on dating the Proto-Anatolian split from residual PIE to the late fifth or early fourth millennium BCE (53, 71) and place the breakup of Anatolian IE inside Turkey before the mid-third millennium (51, 54, 72). In (49) we present new onomastic material (73) 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 before 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 (55). 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 reinterpreted to account for Asia. First, we show that the earliest unambiguous example of horse herding emerged among hunter-gatherers, who had no substantial 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 et al., Population genomics of bronze age Eurasia. *Nature* **522**, 167–172 (2015). doi: [10.1038/nature14507](https://doi.org/10.1038/nature14507)
2. W. Haak et al., Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015). doi: [10.1038/nature14317](https://doi.org/10.1038/nature14317)
3. P. B. Damgaard et al., 137 ancient human genomes from across the Eurasian steppe. *Nature* **557**, 369–374 (2018). doi: [10.1038/s41586-018-0094-2](https://doi.org/10.1038/s41586-018-0094-2)
4. D. W. Anthony, *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2010).
5. A. K. Outram, in *The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers*, V. Cummings, P. Jordan, M. Zvelebil, Eds. (Oxford Univ. Press, Oxford, 2014), pp. 749–766.
6. J. P. Mallory, in *Search of the Indo-Europeans: Language, Archaeology and Myth* (Thames & Hudson, London, 1989).
7. E. R. Jones et al., Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015). doi: [10.1038/ncomms9912](https://doi.org/10.1038/ncomms9912)

8. J. P. Mallory, D. Q. Adams, *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world* (Oxford Univ. Press, Oxford, 2006).
9. K. Kristiansen et al., Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity* **91**, 334–347 (2017). doi: [10.15184/aqy.2017.17](https://doi.org/10.15184/aqy.2017.17)
10. A. K. Outram et al., The earliest horse harnessing and milking. *Science* **323**, 1332–1335 (2009). doi: [10.1126/science.1168594](https://doi.org/10.1126/science.1168594)
11. E. Kaiser, Der Übergang zur Rinderzucht im nördlichen Schwarzmeerraum. *Godišnjak Centar za Balkanološka Ispitivanja* **39**, 23–34 (2010).
12. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* **461**, 489–494 (2009). doi: [10.1038/nature08365](https://doi.org/10.1038/nature08365)
13. K. Kristiansen, *Europe Before History* (Cambridge Univ. Press, Cambridge, 1998).
14. See the supplementary materials.
15. A. K. Outram, A. Polyakov, A. Gromov, V. Moiseyev, A. W. Weber, V. I. Bazaliiskii, O. I. Goriunova, Supplementary discussion of the archaeology of Central Asian and East Asian Neolithic to Bronze Age hunter-gatherers and early pastoralists, including consideration of horse domestication. [10.5281/zenodo.1240521](https://zenodo.org/record/1240521) (9 May 2018).
16. V. V. Evdokimov, V. G. Loman, Raskopki yamnogo kurgana v Karagandinskoi oblasti, in *Voprosy Arheologii Central'nogo i Severnogo Kazakhstana* (Karaganda, 1989), pp. 34–46.
17. D. W. Anthony, D. R. Brown, The secondary products revolution, horse-riding, and mounted warfare. *J. World Prehist.* **24**, 131–160 (2011). doi: [10.1007/s10963-011-9051-9](https://doi.org/10.1007/s10963-011-9051-9)
18. V. I. Bazaliiskii, in *Prehistoric Foragers of the Cis-Baikal, Siberia*, A. Weber, H. McKenzie, Eds. (Northern Hunter-Gatherers Research Series, Canadian Circumpolar Institute Press, Edmonton, 2003), vol. 1.
19. V. I. Bazaliiskii, in *Prehistoric Hunter-Gatherers of the Baikal Region, Siberia: Bioarchaeological Studies of Past Life Ways*, A. Weber, M. A. Katzenberg, T. G. Schurr, Eds. (University of Pennsylvania Press, Philadelphia, 2010).
20. A. Weber, Social evolution among neolithic and early bronze age foragers in the Lake Baikal region: New light on old models. *Arctic Anthropol.* **31**, 1–15 (1994).
21. A. Weber, The Neolithic and Early Bronze Age of the Lake Baikal Region: A review of recent research. *J. World Prehist.* **9**, 99–165 (1995). doi: [10.1007/BF02221004](https://doi.org/10.1007/BF02221004)
22. A. Weber, V. I. Bazaliiskii, “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, 1996), pp. 97–103.
23. V. F. Zaitbert, *Botaiskaya Kultura* (KazAkparat, Almaty, 2009).
24. S. V. Svyatko et al., New radiocarbon dates and a review of the chronology of prehistoric populations from the Minusinsk basin, southern Siberia, Russia. *Radiocarbon* **51**, 243–273 (2009). doi: [10.1017/S0033822200033798](https://doi.org/10.1017/S0033822200033798)
25. V. Siska et al., Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* **3**, e1601877 (2017). doi: [10.1126/sciadv.1601877](https://doi.org/10.1126/sciadv.1601877)
26. M. Raghavan et al., Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014). doi: [10.1038/nature12736](https://doi.org/10.1038/nature12736)
27. Q. Fu et al., The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016). doi: [10.1038/nature17993](https://doi.org/10.1038/nature17993)
28. J. A. Kamm, J. Terhorst, Y. S. Song, Efficient computation of the joint sample frequency spectra for multiple populations. *J. Comput. Graph. Stat.* **26**, 182–194 (2017). doi: [10.1080/10618600.2016.1159212](https://doi.org/10.1080/10618600.2016.1159212)
29. A.-M. Ilumäe et al., Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *Am. J. Hum. Genet.* **99**, 163–173 (2016). doi: [10.1016/j.ajhg.2016.05.025](https://doi.org/10.1016/j.ajhg.2016.05.025)
30. V. T. Kovaleva, N. M. Chirkina, “Etnokul'turnye i etnogeneticheskie protsessy v severnom zaurale v kontse kamennoego-nachale bronzovogo veka: Itogi i problemy issledovaniya” in *Voprosy arheologii Urala* (Uralskii gosudarstvennyi universitet, Ekaterinburg, 1991), pp. 45–70.
31. D. W. Anthony, D. R. Brown, The secondary products revolution, horse-riding, and mounted warfare. *J. World Prehist.* **24**, 131–160 (2011). doi: [10.1007/s10963-011-9051-9](https://doi.org/10.1007/s10963-011-9051-9)
32. C. Gaunitz et al., Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–114 (2018). doi: [10.1126/science.aao3297](https://doi.org/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 *et al.*, A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* **19**, 95–101 (2011). doi: [10.1038/ejhg.2010.146](https://doi.org/10.1038/ejhg.2010.146)
35. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012). doi: [10.1371/journal.pgen.1002453](https://doi.org/10.1371/journal.pgen.1002453)
36. S. Schiffels *et al.*, Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016). doi: [10.1038/ncomms10408](https://doi.org/10.1038/ncomms10408)
37. A. A. Kovalev, D. Erdenebaatar, "Discovery of new cultures of the Bronze Age in Mongolia according to the data obtained by the International Central Asian Archaeological Expedition" in *Current Archaeological Research in Mongolia*, J. Bemmman, H. Parzinger, E. Pohl, D. Tseveendorzh, Eds. (Vor- und Frühgeschichtliche Archäologie Uni-Bonn, Bonn, 2009), pp. 149–170.
38. G. Larson *et al.*, Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8878–8883 (2012). doi: [10.1073/pnas.1203005109](https://doi.org/10.1073/pnas.1203005109)
39. K. H. Røed *et al.*, Genetic analyses reveal independent domestication origins of Eurasian reindeer. *Proc. Biol. Sci.* **275**, 1849–1855 (2008). doi: [10.1098/rspb.2008.0332](https://doi.org/10.1098/rspb.2008.0332)
40. M. A. Zeder, Pathways to animal domestication. *Biodivers. Agric. Domest. Evol. Sustain.* **2012**, 227–259 (2012).
41. R. Willerslev, P. Vitebsky, A. Alekseyev, Sacrifice as the ideal hunt: A cosmological explanation for the origin of reindeer domestication. *J. R. Anthropol. Inst.* **21**, 1–23 (2015). doi: [10.1111/1467-9655.12142](https://doi.org/10.1111/1467-9655.12142)
42. I. Lazaridis *et al.*, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016). doi: [10.1038/nature19310](https://doi.org/10.1038/nature19310)
43. A. Basu, N. Sankar-Roy, P. P. Majumder, Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1594–1599 (2016). doi: [10.1073/pnas.1513197113](https://doi.org/10.1073/pnas.1513197113)
44. H.-P. Francfort, L'âge du bronze en Asie centrale: La civilisation de l'Oxus. *Anthropol. Middle East.* **4**, 91–111 (2009).
45. E. E. Kuz'mina, *The Origin of the Indo-Iranians* (Brill, Leiden, 2007).
46. A. Parpola, *The Roots of Hinduism: The Early Aryans and the Indus Civilization* (Oxford Univ. Press, USA, 2015).
47. I. Mathieson *et al.*, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi: [10.1038/nature16152](https://doi.org/10.1038/nature16152)
48. K. Kristiansen, B. Hemphill, G. Barjamovic, S. Omura, S. Y. Senyurt, V. Moiseyev, A. Gromov, F. E. Yediay, H. Ahmad, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, P. B. Damgaard, Archaeological supplement A to Damgaard *et al.* 2018: Archaeology of the Caucasus, Anatolia, Central and South Asia 4000–1500 BCE. [10.5281/zenodo.1240516](https://doi.org/10.5281/zenodo.1240516) (9 May 2018).
49. G. Kroonen, G. Barjamovic, M. Peyrot, Linguistic supplement to Damgaard *et al.* 2018: Early Indo-European Languages, Anatolian, Tocharian and Indo-Iranian. [10.5281/zenodo.1240524](https://doi.org/10.5281/zenodo.1240524) (9 May 2018).
50. K. Kristiansen, T. B. Larsson, *The Rise of Bronze Age Society: Travels, Transmissions and Transformations* (Cambridge Univ. Press, Cambridge, 2005).
51. H. C. Melchert, "The dialectal position of Anatolian within Indo-European" in *Annual Meeting of the Berkeley Linguistics Society* (1998), pp. 24–31.
52. J. Puhvel, "Whence the Hittite, whither the Jonesian vision?" in *Sprung from a common source*, S. M. Lamb, E. D. Mitchell, Eds. (Stanford Univ. Press, Stanford, Calif., 1991), pp. 51–66.
53. 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.
54. C. Watkins, "An Indo-European linguistic area and its characteristics: Anatolia" in *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, A.Y. Aikhenvald, R. M. W. Dixon, Eds. (Oxford Univ. Press, 2001), pp. 44–63.
55. H. C. Melchert, Ed., *The Luwians* (Brill, 2003).
56. F. Kortlandt, C. C. Uhlenbeck on Indo-European, Uralic and Caucasian. *Hist. Sprachforsch.* **122**, 39–47 (2009).
57. S. M. Winn, Burial evidence and the kurgan culture in Eastern Anatolia c. 3000 BC: An interpretation. *J. Indo-Eur. Stud.* **9**, 113–118 (1981).
58. I. Lazaridis *et al.*, Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
59. 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. Motzenbacher, E. Pernicka, Eds. (Verlag Rudolf Habelt, Bonn, 2010), pp. 263–295.
60. A. T. Smith, *The Political Machine: Assembling Sovereignty in the Bronze Age Caucasus* (Princeton Univ. Press, 2015).
61. A. Sagona, *The Archaeology of the Caucasus: From Earliest Settlements to the Iron Age* (Cambridge Univ. Press, 2017).
62. C. Burney, D. M. Lang, *The Peoples of the Hills: Ancient Ararat and Caucasus* (Weidenfeld and Nicolson, London, 1971).
63. M. T. Larsen, *Ancient Kanesh: A Merchant Colony in Bronze Age Anatolia* (Cambridge Univ. Press, 2015).
64. 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. (Harrassowitz, Wiesbaden, 2010), p. 115.
65. P. Skoglund *et al.*, Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014). doi: [10.1126/science.1253448](https://doi.org/10.1126/science.1253448)
66. N. Lazaridis *et al.*, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi: [10.1038/nature13673](https://doi.org/10.1038/nature13673)
67. M. Mayrhofer, *Die Indo-Arier im Alten Vorderasien: Mit Einer Analytischen Bibliographie* (Harrassowitz, 1966).
68. I. Mathieson *et al.*, The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018). doi: [10.1038/nature25778](https://doi.org/10.1038/nature25778)
69. 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.
70. M. T. Larsen, A. W. Lassen, "Cultural exchange at Kültepe" in *Extraction and Control: Studies In Honor of Matthew W Stolper* (Oriental Institute of the University of Chicago, 2014), pp. 171–188.
71. 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.
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. M. Bonechi, Aleppo in età arcaica. A proposito di un'opera recente. *SEL* **7**, 15–37 (1990).

## ACKNOWLEDGMENTS

We thank K. Magnussen, L. Petersen, and C. Mortensen at the Danish National Sequencing Centre for conducting the sequencing and P. Reimer and S. Hoper at the 14Chrono Center Belfast for providing the AMS dating. We thank S. Ellingvåg, B. E. Heyerdahl, and the Explico-Historical Research Foundation team, as well as N. Thompson, for involvement in field work. We thank the Turkish Ministry of Culture and Tourism, Kaman-Kalehöyük Archeology Museum, and Nevşehir Museum for permission to use samples of Kaman-Kalehöyük and Ovaören. We thank J. Stenderup, P. V. Olsen, and T. Brand for technical assistance in the laboratory. We thank T. Korneliusen for helpful discussions. We thank St. John's College, Cambridge, for providing the setting for fruitful scientific discussions. We thank all involved archaeologists, historians, and collaborators from Pakistan who assisted I.U. in the field. We thank G. Baimbetov (Shejire DNA), I. Baimukhan, B. Daulet, A. Kuseev, A. Koptassarova, Y. Yousupov, M. Akchurin, and V. Volkov for important assistance in the field.

**Funding:** The study was supported by the Lundbeck Foundation (E.W.), the Danish National Research Foundation (E.W.), and KU2016 (E.W.). Research at the Sanger Institute was supported by the Wellcome Trust (grant 206194). R.M. was supported by an EMBO Long-Term Fellowship (ALTF 133-2017). J.K. was supported by the Human Frontiers Science Program (LT000402/2017). Botai fieldwork was supported by University of Exeter, Archeology Exploration Fund, and N. Thompson, Clearwater Documentary. A.B. was supported by NIH grant 5T32GM007197-43. G.K. was funded by Riksbankens Jubileumsfond and European Research Council. M.P. was funded by Netherlands Organization for Scientific Research (NWO), project number 276-70-028. I.U. 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:** E.W., K.K., A.K.O., and A.W. initiated the study. E.W., R.D., K.K., A.K.O., and P.B.D. designed the study. E.W. and R.D. led the study. K.K. and A.K.O. led the archaeological part of the study. G.K., M.P., and G.B. led the linguistic part of the study. P.B.D., C.Z., F.E.Y., I.U., C.d.F., M.I., H.S., A.S.-O., and M.E.A. produced data. P.B.D., R.M., J.K., J.V.M.-M., S.R., K.H.I., M.S., R.N., A.B., J.N., E.W., and R.D. analyzed or assisted in analysis of data. P.B.D., R.M., J.K., J.V.M.-M., R.D., E.W., A.K.O., K.K., G.K., M.P., G.B., B.H., M.S., and R.N. interpreted the results. P.B.D., E.W., R.M., R.D., A.K.O., G.K., J.K., G.B., J.V.M.-M., K.K., and M.P. wrote the manuscript with considerable input from B.H., M.S., M.E.A., and R.N. P.B.D., V.Z., V.M., I.M., N.B., E.U., V.L., F.E.Y., I.U., A.M., K.G.S., V.M., A.G., S.O., S.Y.S., C.M., H.A., A.H., A.S., N.G., M.H.K., A.W., L.O., and A.K.O. excavated, curated, and 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. SNP array data from Pakistan can be obtained from EGA through accession number EGAS00001002965. Y chromosome and mtDNA data are available at Zenodo under DOI [10.5281/zenodo.1219431](https://doi.org/10.5281/zenodo.1219431).

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6396/eaar7711/suppl/DC1](https://www.sciencemag.org/content/360/6396/eaar7711/suppl/DC1)

Supplementary Text

Figs. S1 to S37

Tables S1 to S17

References (74–168)

15 December 2017; accepted 2 May 2018

Published online 9 May 2018

[10.1126/science.aar7711](https://doi.org/10.1126/science.aar7711)



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

Peter de Barros Damgaard, Rui Martiniano, Jack Kamm, J. Víctor Moreno-Mayar, Guus Kroonen, Michaël Peyrot, Gojko Barjamovic, Simon Rasmussen, Claus Zacho, Nurbol Baimukhanov, Victor Zaibert, Victor Merz, Arjun Biddanda, Ilja Merz, Valeriy Loman, Valeriy Evdokimov, Emma Usmanova, Brian Hemphill, Andaine Seguin-Orlando, Fulya Eylem Yediyay, Inam Ullah, Karl-Göran Sjögren, Katrine Højholt Iversen, Jeremy Choin, Constanza de la Fuente, Melissa Ilardo, Hannes Schroeder, Vyacheslav Moiseyev, Andrey Gromov, Andrei Polyakov, Sachihito 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 and Eske Willerslev

*Science* **360** (6396), eaar7711.  
DOI: 10.1126/science.aar7711

### Ancient steppes for human equestrians

The Eurasian steppes reach from the Ukraine in Europe to Mongolia and China. Over the past 5000 years, these flat grasslands were thought to be the route for the ebb and flow of migrant humans, their horses, and their languages. de Barros Damgaard *et al.* probed whole-genome sequences from the remains of 74 individuals found across this region. Although there is evidence for migration into Europe from the steppes, the details of human movements are complex and involve independent acquisitions of horse cultures. Furthermore, it appears that the Indo-European Hittite language derived from Anatolia, not the steppes. The steppe people seem not to have penetrated South Asia. Genetic evidence indicates an independent history involving western Eurasian admixture into ancient South Asian peoples.

*Science*, this issue p. eaar7711

#### ARTICLE TOOLS

<http://science.sciencemag.org/content/360/6396/eaar7711>

#### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/05/08/science.aar7711.DC1>

#### REFERENCES

This article cites 106 articles, 17 of which you can access for free  
<http://science.sciencemag.org/content/360/6396/eaar7711#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)