Article

The genetic origin of the Indo-Europeans

https://doi.org/10.1038/s41586-024-08531-5

Received: 16 April 2024

Accepted: 17 December 2024

Published online: 05 February 2025

Check for updates

Iosif Lazaridis[1,2,67]✉, Nick Patterson[1,3,67]✉, David Anthony[1,4,67]✉, Leonid Vyazov[1,5,67]✉, Romain Fournier[6], Harald Ringbauer[1,7], Iñigo Olalde[1,8,9], Alexander A. Khokhlov[10], Egor P. Kitov[11], Natalia I. Shishlina[12], Sorin C. Ailincăi[13], Danila S. Agapov[14], Sergey A. Agapov[14], Elena Batieva[15], Baitanayev Bauyrzhan[16], Zsolt Bereczki[17], Alexandra Buzhilova[18], Piya Changmai[5], Andrey A. Chizhevsky[19], Ion Ciobanu[20], Mihai Constantinescu[21], Marietta Csányi[22], János Dani[23,24], Peter K. Dashkovskiy[25], Sándor Évinger[26], Anatoly Faifert[27], Pavel Flegontov[1,5,28], Alin Frînculeasa[29], Mădălina N. Frînculeasa[30], Tamás Hajdu[31], Tom Higham[32,33], Paweł Jarosz[34], Pavol Jelínek[35], Valeri I. Khartanovich[36], Eduard N. Kirginekov[37], Viktória Kiss[38], Alexandera Kitova[39], Alexeiy V. Kiyashko[40], Jovan Koledin[41], Arkady Korolev[10], Pavel Kosintsev[42,43], Gabriella Kulcsár[38], Pavel Kuznetsov[10], Rabadan Magomedov[44], Aslan M. Mamedov[16], Eszter Melis[38], Vyacheslav Moiseyev[36], Erika Molnár[17], Janet Monge[45], Octav Negrea[29], Nadezhda A. Nikolaeva[46], Mario Novak[47,48], Maria Ochir-Goryaeva[49], György Pálfi[17], Sergiu Popovici[50], Marina P. Rykun[51], Tatyana M. Savenkova[52], Vladimir P. Semibratov[53], Nikolai N. Seregin[54], Alena Šefčáková[55], Raikhan S. Mussayeva[16], Irina Shingiray[56], Vladimir N. Shirokov[57], Angela Simalcsik[20,58], Kendra Sirak[1,2], Konstantin N. Solodovnikov[59], Judit Tárnoki[22], Alexey A. Tishkin[53], Viktor Trifonov[60], Sergey Vasilyev[61], Ali Akbari[1,2], Esther S. Brielle[1], Kim Callan[2,62], Francesca Candilio[63], Olivia Cheronet[32,33], Elizabeth Curtis[2,62], Olga Flegontova[5,28], Lora Iliev[2,62], Aisling Kearns[2], Denise Keating[64], Ann Marie Lawson[2,62], Matthew Mah[2,3,62], Adam Micco[2], Megan Michel[1,2,7,62], Jonas Oppenheimer[2,62], Lijun Qiu[2,62], J. Noah Workman[2,62], Fatma Zalzala[2,62], Anna Szécsényi-Nagy[65], Pier Francesco Palamara[6,66], Swapan Mallick[2,3,62], Nadin Rohland[2], Ron Pinhasi[32,33]✉ & David Reich[1,2,3,62]✉

The Yamnaya archaeological complex appeared around 3300 BC across the steppes north of the Black and Caspian Seas, and by 3000 BC it reached its maximal extent, ranging from Hungary in the west to Kazakhstan in the east. To localize Yamnaya origins among the preceding Eneolithic people, we assembled ancient DNA from 435 individuals, demonstrating three genetic clines. A Caucasus–lower Volga (CLV) cline suffused with Caucasus hunter-gatherer[1] ancestry extended between a Caucasus Neolithic southern end and a northern end at Berezhnovka along the lower Volga river. Bidirectional gene flow created intermediate populations, such as the north Caucasus Maikop people, and those at Remontnoye on the steppe. The Volga cline was formed as CLV people mixed with upriver populations of Eastern hunter-gatherer[2] ancestry, creating hypervariable groups, including one at Khvalynsk. The Dnipro cline was formed when CLV people moved west, mixing with people with Ukraine Neolithic hunter-gatherer ancestry[3] along the Dnipro and Don rivers to establish Serednii Stih groups, from whom Yamnaya ancestors formed around 4000 BC and grew rapidly after 3750–3350 BC. The CLV people contributed around four-fifths of the ancestry of the Yamnaya and, entering Anatolia, probably from the east, at least one-tenth of the ancestry of Bronze Age central Anatolians, who spoke Hittite[4,5]. We therefore propose that the final unity of the speakers of 'proto-Indo-Anatolian', the language ancestral to both Anatolian and Indo-European people, occurred in CLV people some time between 4400 BC and 4000 BC.

Between 3300 BC and 1500 BC, people of the Yamnaya archaeological complex and their descendants spread Indo-European languages from the steppe[2,6–12] and transformed Europe, Central and South Asia, Siberia and the Caucasus. Sparse sampling of Yamnaya people and their Eneolithic precursors creates a problem for understanding the origins of this Bronze Age culture. It is broadly accepted that the Yamnaya had two ancestries: northern, eastern hunter-gatherer (EHG) ancestry from far-eastern Europe, and southern, West Asian ancestry[2] from Caucasus hunter-gatherers (CHG) in Georgia[1] and Neolithic people from Zagros[13] and the south Caucasus[10,14,15]. These two groups interacted across West Asia and eastern Europe[13], but it has not been clear where or how the Eneolithic ancestors of the Yamnaya first appeared. Potential northern ancestors include the EHG, and EHG mixed with western hunter-gatherers[16] (WHG), for example in the Dnipro

A list of affiliations appears at the end of the paper.

valley[3], where they formed the Ukraine Neolithic hunter-gatherers (UNHG). But the Yamnaya also received Anatolian Neolithic ancestry[9], mediated by Caucasus Neolithic populations, such as those sampled at Aknashen and Masis Blur in Armenia[10], and even possibly Siberian ancestry that reached the European steppe before their emergence[9].

We present a genetic analysis of 367 newly reported individuals (6400–2000 BC) and increased data quality for 68 individuals[6] (a total of 435 individuals). The present study is the formal report for 291 and 63 of these, respectively; more than 80% are from Russia, and the rest are largely from the western expansion into the Danube valley (Supplementary Information section 1 and Supplementary Table 1). Details of 803 ancient DNA libraries (195 that failed screening) are in Supplementary Information section 1 and Supplementary Table 2, and 198 new radiocarbon dates are in Supplementary Table 3. A parallel study[17] of the North Pontic Region (Ukraine and Moldova) is the formal report for the remaining individuals. We labelled individuals on the basis of geographical and temporal information, archaeological context and genetic clustering (Supplementary Information section 1 and Supplementary Table 4). The combined dataset adds 79 Eneolithic people from the European steppe and its environs to 82 published. It also adds 211 Yamnaya (and related Afanasievo) individuals to the 75 previously published (Methods).

Three pre-Bronze Age genetic clines

Principal component analysis (PCA) of ancient individuals from the Pontic–Caspian steppe and adjacent areas reveals that Eneolithic people and the Bronze Age Yamnaya fall on non-overlapping gradients (Fig. 1 and Supplementary Table 5). PC1 correlates (right to left) to differentiation between inland West Asian (Caucasus and Iran) and East Mediterranean populations (Anatolian–European)[14], but interpretation is not clear because this axis also correlates to differentiation between Siberian and European hunter-gatherers. PC2 differentiates between northern Eurasians (top, including Europe and Siberia) and West Asians (bottom, Anatolia–Mesopotamia–Caucasus–Iran). Eneolithic and Bronze Age people occupy the middle, indicating that they formed by mixture.

To distinguish alternative mixture scenarios that could explain these patterns, we implemented a competition framework around qpWave/qpAdm[2,18] (Methods and Supplementary Information section 2). The idea is that model *X* (a set of admixing sources) describes a target population *T* if: it reconstructs the shared genetic drift of *T* with both distant outgroup populations and the sources of alternative models; and also renders these models infeasible if they cannot model shared drift with the sources of *X*. Models are thus first filtered against a set of distant outgroups; having survived this step, they are compared all-against-all to produce a set of promising models.

Three PCA clines (denoted geographically as Volga, Dnipro and Caucasus–lower Volga) diverge from the area enclosed by the Lower Don (at Krivyansky), lower Volga (at Berezhnovka-2) and north Caucasus (at Progress-2, Vonyuchka-1 and Sharakhalsun[9]). They extend from there towards: EHG and UNHG, representing the pre-Eneolithic people of the Volga–Don–Dnipro area of eastern Europe; and CHG and Caucasus Neolithic, representing the pre-Eneolithic people of the Caucasus and West Asia.

The Volga cline

Distinct upriver and downriver gradients formed by Eneolithic individuals who lived on waterways that drain into the Caspian Sea delineate zones of ongoing human contact. PCA positions correlate well to positions along the Volga: the Volosovo-attributed Sakhtysh (in the upper Volga) and Murzikha (near the Kama–Volga confluence)[19] constitute the upriver European hunter-gatherer cline, between EHG and UNHG. A 'bend' separates the two clines and is occupied by EHG groups, including middle Volga ones and those from northwest Russia in Karelia[2,20], which is a very wide geographic distribution indicating that EHG was the earlier established population. Downriver and past the bend, we find the Volga cline: hunter-gatherer affinity decreases at the middle Volga at Labazy, Lebyazhinka, Ekaterinovka, Syezzheye then Khvalynsk (4500–4350 BC) and Khlopkov Bugor, before reaching the lower Volga at Berezhnovka-2 (4450–3960 BC) (Fig. 1a,b). This decrease is counterbalanced by increased affinity to the Caucasus, driven by an unsampled CHG-related source, somewhere between Georgia (the sampling location of CHG[1]) and the lower Volga, interacting with EHG people. Archaeological correlates for such interactions begin with the expansion of the Seroglazovo forager culture around the lower Volga estuary in around 6200 BC, which parallels cultures of the Caucasus in ceramics and lithics, and continue to the north Caucasus Neolithic cemetery near Nalchik, dated to around 4800 BC[21,22].

At the end of the Volga cline, four lower-Volga individuals from Berezhnovka-2 can be grouped with the north Caucasus PG2004 individual from Progress-2 (ref. 9), dated to 4240–4047 BC, into a Berezhnovka 2–Progress-2 cluster labelled the BPgroup. The second Progress-2 individual (PG2001; 4994–4802 cal BC) groups with another north Caucasus individual from Vonyuchka-1 (ref. 9; VJ1001; 4337–4177 BC) into a Progress-2–Vonyuchka 1 cluster (the PVgroup). The BPgroup and PVgroup are distinct ($P = 0.0006$) but little differentiated (fixation index $F_{ST} = -0.002 \pm 0.002$; Extended Data Table 1), indicating movement between the north Caucasus piedmont and the lower Volga. These two locations also shared a distinctive burial pose, on the back with raised knees, which was later typical of the Yamnaya and dated earliest in four individuals from Ekaterinovka (4800–4500 BC), contrasting with 95% of the graves, which had individuals posed supine with legs extended straight, and also a female (individual 2) from Lebyazhinka-5, grave 12 (4838–4612 BC). BPgroup is shifted relative to PVgroup (Fig. 1b) towards Afontova Gora-3 from Upper Palaeolithic Siberia[23], West Siberian hunter-gatherers[8] and a Neolithic individual dated at 7,500 years ago from Tutkaul (TTK) from Central Asia[20].

A natural interpretation is that upriver, EHG-related, and downriver, Berezhnovka-related, ancestors came together along the Volga, forming the genetic gradient. The upriver ancestry has long-established eastern European antecedents[20], unlike the downriver ancestry, because: first, there are no earlier sequenced individuals from the lower Volga; second, the Berezhnovka people are distinct from preceding groups; and third, BPgroup cannot be modelled as a clade with contemporary or earlier groups ($P < 0.001$). Whatever BPgroup's origins are, we can use it as one proximate source for the Volga cline together with an EHG source from Karelia[2,20], which is well outside the Volga area and is thus unlikely to be part of the riverine mating network. Seven Volga cline populations fit this model (*P*-values of 0.04 for Ekaterinovka and 0.12–0.72 for the others) with consistently poor fits only for upper Volga, Murzikha, Maximovka and Klo (the Khvalynsk individuals with low Berezhnovka relatedness) (*P*-values from 1×10^{-66} to 0.006). Three of these (other than Klo) are arrayed in the upriver EHG cline (Fig. 1c).

People buried at Ekaterinovka (5050–4450 BC, based on three herbivore bone radiocarbon dates unaffected by marine reservoir effects; Supplementary Table 1) were already mixing with lower Volga Berezhnovka-related people (24.3 ± 1.3%). This contrasts with the earlier hunter-gatherers from Lebyazhinka (7.9 ± 3.6%; consistent with zero, $P = 0.21$). A century or two later at Khvalynsk[24], around 120 km from Ekaterinovka (4500–4350 BC, based on two herbivore bones), there is an admixture gradient, divided for convenience into: Khvalynsk high (Khi; 76.8 ± 1.9% BPgroup), Khvalynsk medium (Kmed; 57.3 ± 1.7% BPgroup) and Khvalynsk low (Klo; 41.2 ± 1.6% BPgroup). Volga cline individuals had around 14–89% Berezhnovka ancestry (Fig. 1c), dominated by neither the old native EHG group nor the lower Volga newcomers. Genetic differentiation between lower Volga (BPgroup) and Ekaterinovka was

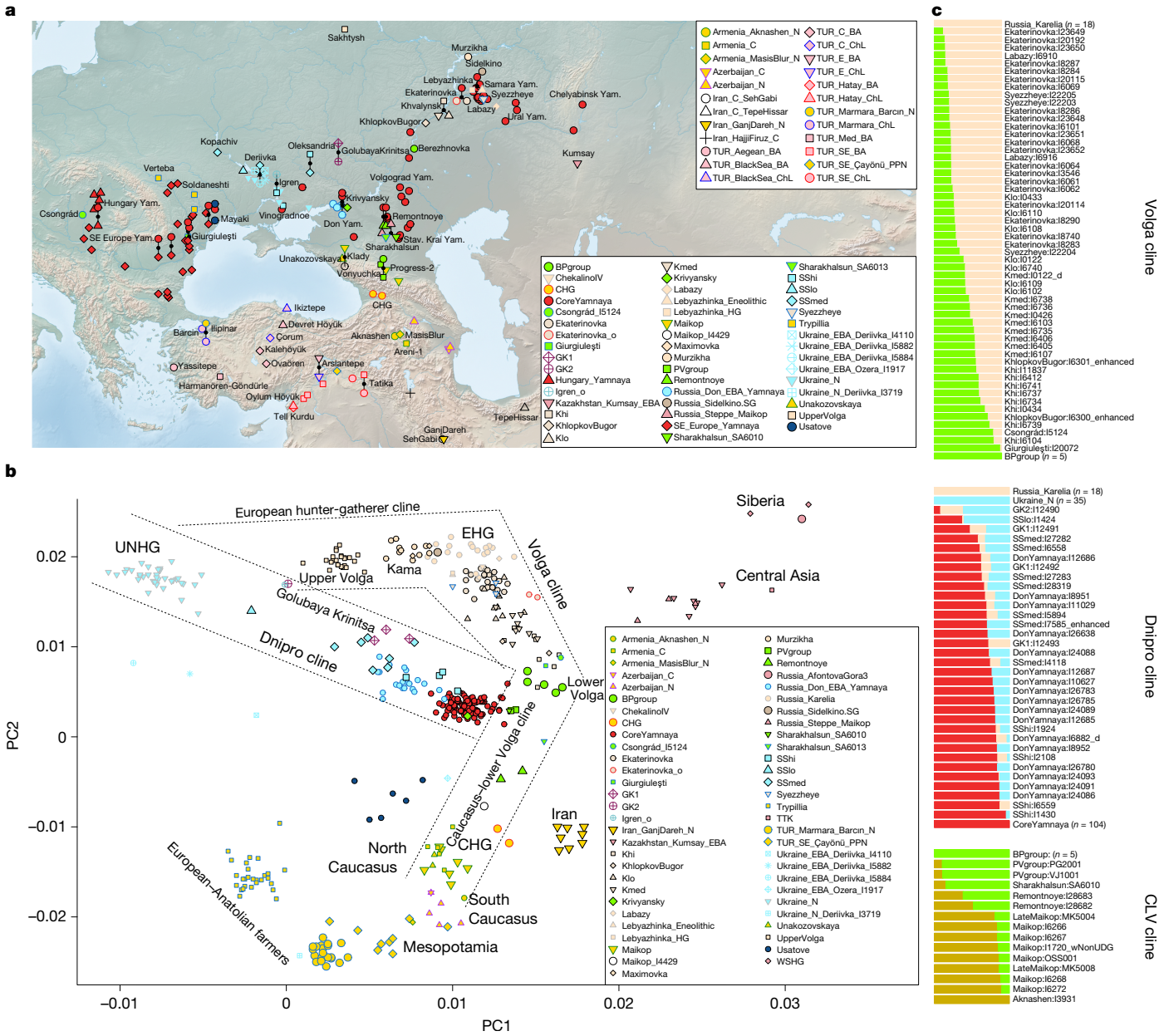


Fig. 1 | Three Eneolithic clines and their neighbours in space and time. **a**, Map with analysed sites. **b**, PCA using axes formed by a set of ancient West European hunter-gatherer and Siberian, West Asian and European farmer populations. Selected individuals relevant to this study are projected (Methods). **c**, qpAdm models fitted on individuals of the populations of the clines. The Volga cline is generated by admixture between lower Volga (BPgroup) people with upriver EHG populations. People of the Dnipro cline have UNHG or UNHG + EHG admixture relative to the Core Yamnaya (the hunter-gatherer source along this cline is significantly variable). The Caucasus–lower Volga cline is generated by admixture of lower Volga people with those from the Neolithic Caucasus (Aknashen related). The map was drawn using public-domain Natural Earth data with the rnaturalearth package in R[54].

strong (F_{ST} = 0.030 ± 0.001; Extended Data Table 1), probably reflecting different linguistic–cultural communities.

A genetically Volga cline individual from Csongrád-Kettőshalom in Hungary (4331–4073 BC) had 87.9 ± 3.5% BPgroup ancestry (Fig. 1c), similar to Khi individuals. This individual was from late fifth millennium BC steppe-like graves in southeastern Europe that included a cemetery at Mayaky in Ukraine[17,25,26] and a cemetery at Giurgiuleşti[27] in Moldova, from which one individual (I20072; 4330–4058 BC) is a clade with BPgroup (P = 0.90). Archaeology has documented Balkan copper on the Volga cline site of Khvalynsk[24], and the Csongrád and Giurgiuleşti individuals were plausibly part of this cultural exchange, leapfrogging the intervening Dnipro and Don basins without picking up ancestry from them[17].

The Dnipro cline

The Dnipro cline is formed by Neolithic individuals who lived along the Dnipro River rapids (UNHG; 6242-4542 BC) and the Serednii Stih population, represented by 13 individuals (4996–3372 BC; uncorrected for freshwater-reservoir effects). This cline also includes most later Yamnaya individuals, a high-quality and genetically homogeneous subset (n = 104) that we term Core Yamnaya (Supplementary Information section 2). Close to Core Yamnaya (Fig. 1b) are some Eneolithic individuals: the Serednii Stih individual from Krivyansky in the lower Don (4359–4251 BC) and the PVgroup from the north Caucasus. Nonetheless, the Core Yamnaya cannot be modelled as derived from them or any other single source ($P < 1 \times 10^{-4}$). Dnipro cline people are also

distinct from Volga cline individuals because no inter-riverine pairs form a clade ($P < 1 \times 10^{-7}$). This distinctiveness spans three millennia, commencing with the UNHG, continuing with the Eneolithic Serednii Stih, and ending with the Early Bronze Age Yamnaya. A geographically localized Yamnaya population of the lower Don ($n = 23$), many ($n = 17$) from the site of Krivyansky, is distinct from the Eneolithic individual at Krivyansky (Fig. 1b) and not a clade with them ($P = 8 \times 10^{-15}$). The Yamnaya can thus not be traced to the north Caucasus (PVgroup), the lower Don (Krivyansky) or the Volga (BPgroup and the rest of the Volga cline). Their placement on the Dnipro cline indicates their formation by a process of admixture as descendants of the Serednii Stih culture.

Serednii Stih heterogeneity contrasts with Core Yamnaya homogeneity (Fig. 1b), which is remarkable given the 5,000-km-wide sampling of the latter, from Hungary to southern Siberia. The Yamnaya expanded across this vast region, hardly admixing with locals, at least initially and for the elite individuals buried in kurgans. Individuals of the Serednii Stih culture are arrayed along the Dnipro cline. An individual from Vinogradnoe, grouped with two from Oleksandria and one from Igren, fall into an SShi cluster of greatest Core Yamnaya affinity but are not a clade with them ($P = 2 \times 10^{-7}$). A Kopachiv female (I7585)[26] is part of an SSmed cluster further along the cline, which also includes three individuals from Oleksandria and three from Deriivka. SShi and SSmed are largely contiguous, but I1424 from Moliukhiv Bugor (SSlo) is apart from them, close to UNHG. Variation within the Serednii Stih plausibly included unsampled individuals in gaps along the cline, or beyond its sampled variation. The Don Yamnaya largely overlap with the Serednii Stih, and at stratified sites of the lower Don Konstantinovka culture, they continued to occupy Serednii Stih settlements, a continuity unobserved in the Volga–Ural steppes.

All Dnipro cline groups can be well modelled with either UNHG or GK2 (individual I12490 from Golubaya Krinitsa in the middle Don; 5610–5390 BC) at one extreme, and Core Yamnaya on the other (*P*-values 0.07–0.85). However, the hunter-gatherer end of the cline is not clearly one or the other; although the source for SSmed upriver fits just as well as UNHG ($P = 0.27$) or GK2 ($P = 0.43$), the Don Yamnaya upriver source can fit only as UNHG ($P = 0.08$), not GK2 ($P = 0.0001$), and the SShi upriver source can fit only as GK2 ($P = 0.08$), not UNHG ($P = 0.003$). We therefore model individuals from any point along the entire UNHG–EHG cline (Fig. 1c), not presupposing either UNHG or GK2 as the source, finding that UNHG ancestry predominates but more EHG ancestry is also present (as at GK2). The hunter-gatherer source was thus from the Dnipro–Don (UNHG–GK2), not the Volga (EHG). GK2 clusters with Mesolithic hunter-gatherers from Vasylivka in the Dnipro[17] and may stand in for unsampled survivors there of that earlier population. Core Yamnaya as a source for earlier populations would be ahistorical; it must stand for an unsampled Eneolithic source.

The Don, which lies between the Dnipro and the Volga, is represented by middle Don Golubaya Krinitsa individuals and the lower Don Krivyansky. Golubaya Krinitsa contained archaeologically contrasting graves, one similar to those of the Dnipro Neolithic and the other similar to Serednii Stih[28]. GK2 is modelled as 66.6 ± 4.7% UNHG and 33.4 ± 4.7% EHG ($P = 0.39$). Using the most ancient sources (Karelia, UNHG and CHG), Krivyansky Eneolithic and Golubaya Krinitsa individuals have variable CHG-related ancestry (Fig. 2a), maximized at Krivyansky (58.9 ± 2.4%) and less (25.3 ± 2.1%) in three Golubaya Krinitsa individuals grouped as GK1 (Fig. 1); GK2 had none or little (4.0 ± 2.2%). Thus, the admixture history of the Don paralleled its intermediate geography, and included southern, CHG-related ancestry (Fig. 2a). This was already present in GK1 (individual I12491; 5557–5381 BC)[11], indicative of an early presence, but its absence in GK2 of a similar age shows that it was not generally present. Dates for GK1 and GK2 may be inflated because Golubaya Krinitsa was archaeologically interpreted as being in cultural contact with the much later Eneolithic Serednii Stih[29]. Moreover, a Serednii Stih outlier from Igren (I27930; 4337–4063 cal BC) is a clade with GK2; this could be evidence of long-distance migration from the Don to the Dnipro in a Serednii Stih time frame. ^{14}C dates at Golubaya Krinitsa could potentially be overestimated owing to the consumption of freshwater fish, which inflate dates by up to a millennium in this region[30].

It has been suggested[11] that the Yamnaya had roughly 35% CHG-related and about 65% Golubaya Krinitsa ancestry, the latter already having around 20–30% CHG-related ancestry, implying that the main Yamnaya source may have been hunter-gatherers of the Don area. Contradicting this model, Yamnaya do not fit models with CHG-related and either GK1 or GK2 sources[11] ($P < 10^{-6}$). To better understand this, we fitted Yamnaya to a model of Karelia + UNHG + CHG (Fig. 2a) and found that it underestimates the shared drift of Core Yamnaya with both Afontova Gora-3 from Upper Palaeolithic Siberia ($Z = -5.2$) and Anatolian Neolithic ($Z = -6.8$). A Volga source of the Siberian-related ancestry is indicated by the fact that applying the same model to Volga cline groups also underestimates shared drift with Afontova Gora-3 ($P = 1 \times 10^{-8}$ and $Z = -4.5$ for BPgroup; the Siberian ancestry is also evident in the deviation of the Dnipro cline towards Siberians in Fig. 1b). This Siberian-related ancestry is also affirmed because BPgroup can be modelled as around 76% Krivyansky and 24% Central Asian (Siberian related) Tutkaul[20] ($P = 0.13$). When we fit Krivyansky and BPgroup with the model that includes all relevant ancestries, CHG, GK2 and Tutkaul (Fig. 2b), Krivyansky has little to no Central Asian ancestry (5.1 ± 3.6%), fitting as a simple two-way mix of 56.7 ± 2.6% CHG related and 43.3 ± 2.6% GK2 ($P = 0.37$). By contrast, BPgroup requires 29.3 ± 2.2% Tutkaul. Even adding Siberian-related ancestry (Tutkaul) is not sufficient to model the Core Yamnaya, however, because the three-way model in Fig. 2b still fails ($P = 10^{-9}$) to explain the shared drift with Anatolian Neolithic ($Z = -6.1$).

Central Asian or Siberian ancestry was therefore already in the north Caucasus steppe and Volga during the Neolithic, but with no evidence of it further west on the Don. Adding a third, western (UNHG) or eastern (Tutkaul), source (Fig. 2c,d) to the two-source BPgroup + EHG model for Volga cline individuals, they remain well modelled with these two alone (Fig. 2c). Some have more Tutkaul ancestry (Fig. 2d). However, deviations are minor (4.4 ± 2.6% Tutkaul ancestry for Khi). Crucially, the Core Yamnaya fail all models of Fig. 2a–d ($P < 10^{-8}$), so they were not formed from the CHG–EHG–UNHG–Tutkaul blend of these models.

The CLV cline

The Core Yamnaya, positioned on the opposite end of the Dnipro cline to the UNHG and GK2 (Fig. 1b), had ancestry from an unknown source of lower or even no such ancestry. The only consistently fitting ($P = 0.67$) two-way model for them involved 73.7 ± 3.4% of the SShi subset of Serednii Stih and 26.3 ± 3.4% from a population represented by two Eneolithic individuals from Sukhaya Termista I (I28682) and Ulan IV (I28683) (4152–3637BC) near the village of Remontnoye, north of the Manych Depression between the lower Don and the Caspian Sea. Remontnoye is on neither the Volga nor the Dnipro cline and does not form a clade ($P < 10^{-10}$) with any other group. It had at least two sources: a southern, Caucasus one, comprising either descendants of people like those who lived in Neolithic Armenia at Aknashen[10], or ancestors of people of the Bronze Age north Caucasus Maikop[9] culture; and a northern one, from a population like BPgroup. The southern component can be modelled as having around half its ancestry from either Aknashen (44.6 ± 2.7%; $P = 0.66$) or Maikop (48.1 ± 2.9%; $P = 0.44$). We estimate −0.3 ± 2.9% UNHG or −0.5 ± 3.5% GK2 ancestry when either is added as a third source to the Aknashen + BPgroup model, so Remontnoye had no discernible UNHG/GK2-related ancestry as anticipated for the unknown source for the Yamnaya. Moreover, the main Maikop cluster, including individuals buried in kurgans in Klady and Dlinnaya-Polyana, had 86.2 ± 2.9% ($P = 0.50$) Aknashen ancestry. Thus, there is a CLV cline: Aknashen–Maikop–Remontnoye–Berezhnovka. These four, arrayed in order of decreasing Caucasus Neolithic component, match their south-to-north location. North Caucasus people

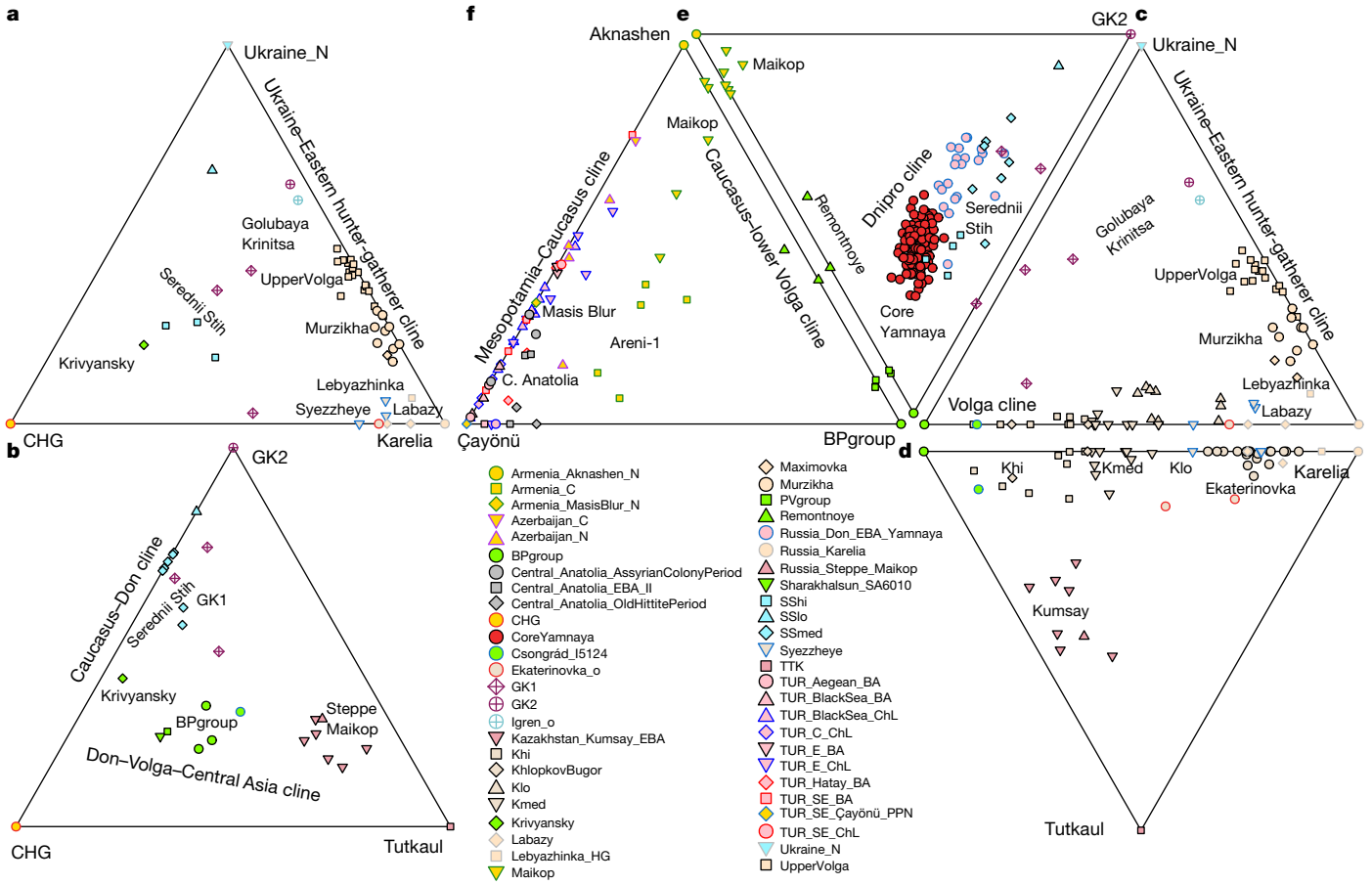


Fig. 2 | The three clines in the context of Eneolithic and Bronze Age admixture. Six three-source qpAdm models elucidate a complex history of admixture. **a**, Caucasus and European hunter-gatherer admixtures in the 'old steppe': Krivyansky on the lower Don received much more CHG-related admixture than did upriver people of the middle Don at Golubaya Krinitsa. In the middle and upper Volga and the Kama River, populations had negligible CHG-related influence. **b**, The Don–Volga difference. On the lower Volga and North Caucasus piedmont, the BPgroup received CHG-related ancestry, similar to its western lower Don counterpart at Krivyansky. But it also received ancestry from Central Asia, and this eastern influence was greater still in the Bronze Age steppe Maikop. **c**, The Volga basin Eneolithic populations with regard to the Don. Populations at Khvalynsk, Klopkov Bugor and Ekaterinovka form a Volga cline between the Berezhnvoka cluster on the lower Volga and the upriver EHG-like populations of the middle Volga (Labazy and Lebyazhinka). **d**, The Volga basin Eneolithic populations with regard to Central Asia. There is a slight excess of Central Asian ancestry in the Khi subset of Khvalynsk. **e**, The Dnipro cline. The Core Yamnaya are on one end of a cline that also includes the Don Yamnaya and Seredнii Stih populations, formed by admixture of North Pontic hunter-gatherers with those from the CLV cline of differential admixture of Neolithic Caucasus and BPgroup people. The CLV cline includes diverse people buried in kurgans at Berezhnovka, Progress-2, Remontnoye and the Maikop sites of Klady and Dlinnaya-Polyana, dating from around 5000–3000 BC. **f**, West Asia. CLV ancestry first appears in the Chalcolithic population at Areni-1 in Armenia and is also present in the Bronze Age at Maikop. Most of the ancestry is from West Asian sources from the Mesopotamia–Caucasus (or Çayönü–Masis Blur–Aknashen) cline. Chalcolithic and Bronze Age Anatolians lack CLV ancestry, but traces of it can be found in Bronze Age central Anatolians.

at Progress-2 and Vonyuchka-1 bucked the latitudinal trend, having, unlike their Maikop neighbours, little Caucasus Neolithic ancestry. These violations document long-range connectivity across the CLV area, and provide an important example of how genetics and geography do not always match.

We wanted to know which group mediated the southern ancestry of the CLV cline. It is not Aknashen, which is geographically remote and much earlier (5985–5836 BC). It is not Maikop, which was geographically closer but later (3932–2934 BC). Unsampled Meshoko and Svobodnoe settlements (4466–3810 BC)[31] are plausible for the expansion of Aknashen-like ancestry northward and Berezhnovka-like ancestry southward, because they exchanged exotic stone, copper and stone mace heads with Volga cline sites. They are preceded in the north Caucasus by the Eneolithic Unakozovskaya (ref. 9, 4607–4450 BC, and this study) and succeeded by the Maikop. The Unakozovskaya population is not a good genetic source for Remontnoye, because the model BPgroup + Unakozovskaya fails ($P < 0.001$) by overestimating ($Z = 3.8$) CHG-related drift. Unakozovskaya is well modelled as 95.3 ± 6.3% Maikop and 4.7 ± 6.3% CHG ($P = 0.46$); this group is therefore Maikop-like, but distinct genetically ($P = 2 \times 10^{-11}$) (Fig. 1b). A recently published[32] individual from Nalchik (around 5000–4800 cal BC) had more steppe affinity than the sampled Unakozovskaya, and can be modelled (Supplementary Information section 2) as a mix of Unakozovskaya and steppe populations. Thus, in the Eneolithic north Caucasus there were: Aknashen-related ancestry, representing the Neolithic spread; CHG-related ancestry, indicated by the Maikop–Unakozovskaya contrast; and northern lower Volga ancestry, constituting about one-seventh of the ancestry of the sampled Maikop.

Remontnoye, Berezhnovka and Maikop all used kurgan burial, which was common at around 5000–3000 BC in diverse CLV cline people. By contrast, a distinctive burial feature, with individuals posed on the back with the knees raised and the floor of the burial pit covered with red ochre, was shared by almost all steppe groups including the Seredнii Stih and Volga cline, while Remontnoye and Maikop burials were contracted on one side. Some funeral customs united Maikop with the steppes, but others separated them.

The CLV cline reveals that the ancestors of Dnipro cline Serednii Stih and Yamnaya were CLV cline people, similar to Remontnoye, who had moved into the Dnipro–Don region and mixed with locals. The actual sources for the Yamnaya may have differed from the sampled Remontnoye and SShi. The Dnipro cline can be fit (Fig. 2e) by a three-way model in which a Dnipro or Don hunter-gatherer source mixed with groups of mixed Aknashen and Berezhnovka ancestry. Either GK2 or UNHG can fit as the northern riverine source, but we use GK2 in Fig. 2e because this model has a higher *P*-value (0.93) than the UNHG alternative ($P = 0.04$). The Yamnaya are inferred to have about one-fifth of their ancestry from Dnipro/Don hunter-gatherers: either 22.5 ± 1.8% GK2 or 17.7 ± 1.3% UNHG.

The CLV cline was the source from which Caucasus-derived ancestry flowed into the ancestors of the Yamnaya[10]. The Remontnoye + SShi model predicts shared genetic drift with Neolithic Anatolians well ($Z = -0.8$), unlike models lacking Anatolian Neolithic ancestry (Fig. 2a–d). Archaeology has established that trade in Balkan copper during the late fifth millennium BC to north Caucasus farmer sites (Svobodnoe) and the Volga (Khvalynsk) took place, and Neolithic pots similar to those from Svobodnoe appeared in Dnipro–Don steppe sites connected with the Seredni Stih culture (Novodanilovka). This cultural exchange contextualizes the entry of BPgroup/Aknashen mixed groups into the Dnipro–Don steppes.

CLV impact in Armenia and Anatolia

People of the CLV cline also went south (Fig. 2f), explaining the steppe ancestry found at Areni-1 in Chalcolithic Armenia from around 4000 BC[13], where lower Volga ancestry (26.9 ± 2.3% BPgroup) admixed with a local Masis Blur-related Neolithic substratum (Supplementary Information section 2). This contrasts with the north Caucasus Maikop, where the substratum was Aknashen related. We can model Masis Blur as 33.9 ± 8.6% Aknashen and 66.1 ± 8.6% Pre-Pottery Neolithic of the Tigris Basin of Mesopotamia[33] at Çayönü ($P = 0.47$), part of a Neolithic Çayönü–Masis Blur–Aknashen cline. The populations of Armenia retained CHG differentially[6]: more (42.0 ± 3.8%) in Aknashen than in Masis Blur (13.7 ± 4.0%). Some Anatolian Chalcolithic and Bronze Age groups can be derived entirely from the Caucasus–Mesopotamian cline (Fig. 2f), whereas others also have ancestry from the Mesopotamian–Anatolian cline, lacking any steppe ancestry[10,15,34–36].

We show that Central Anatolians[34] from the Early Bronze Age (2750–2500 BC), Assyrian Colony (2000–1750 BC) and Old Hittite (1750–1500 BC) periods were unusual in the Anatolian landscape because they had CLV ancestry combined with Mesopotamian (Çayönü) (Fig. 2f, Extended Data Fig. 1 and Supplementary Information section 2). The non-Mesopotamian ancestry varied, depending on the level of CLV input: 10.8 ± 1.7% ancestry ($P = 0.14$) from BPgroup, 19.0 ± 2.4% from Remontnoye ($P = 0.19$) or 33.5 ± 4.8% from Armenia_C ($P = 0.10$).

The exact source of the steppe ancestry in Anatolia cannot be precisely determined, but all fitting models involve some of it (Extended Data Fig. 1a). Some of the steppe-related sources are unlikely on chronological or linguistic grounds; for example, the Core Yamnaya (12.2 ± 2.0%; $P = 0.10$), as well as western Yamnaya-derived populations from southeastern Europe, such as from Boyanovo or Mayaky Early Bronze Age[25] (Extended Data Fig. 1b). The Early Bronze Age Central Anatolians from Ovaören[34] (2750–2500 BC) do temporally overlap the late Yamnaya period, but the timing of the Yamnaya expansion is in tension with the much-earlier linguistic split of Anatolian languages that form an outgroup to those of the inner Indo-European Core[37]. Fixing Çayönü as one source and adding pairs of steppe sources (allowing ancestry to range freely along the Volga, Dnipro and CLV clines), the hunter-gatherer contribution is negative on the Volga cline (−3.4 ± 2.6% EHG) and on the Dnipro cline (−2.3 ± 2.7% UNHG and −3.9 ± 3.5% GK2); thus, the admixing population had no more EHG, UNHG or GK2 ancestry than did the BPgroup or Core Yamnaya endpoints of these two clines (Supplementary Information section 2). Placing the admixing population on the CLV cline is successful ($P = 0.129$), with a significant amount of BPgroup ancestry (8.8 ± 2.7%) validating a CLV and north-of-the-Caucasus mountains Eneolithic origin. Steppe + Mesopotamian models fit the Central Anatolian Bronze Age but none of the Chalcolithic/Bronze Age Anatolian regional subsets ($P < 0.001$; the BPgroup + Çayönü model is shown in Extended Data Fig. 1c): their success is not due to their general applicability. Moreover, steppe ancestry in the Central Anatolian Bronze Age is observed across individuals and periods (Extended Data Fig. 1d), including Early Bronze Age Ovaören south of the Kızılırmak river and Middle or Late Bronze Age Kalehöyük just within the bend of the river[34]. This is consistent with an Anatolian–Hattic linguistic boundary coinciding with the Kızılırmak, a boundary breached before the conquest of Hattusa by the Hittites in roughly 1730 BC[4]. Regardless of the (inherently unknowable) linguistic identity of the sampled individuals, their unique blend of ancestries demands an explanation.

Populations along the path to Central Anatolia can be modelled with BPgroup ancestry and distinctive Caucasus–Mesopotamian substrata: Aknashen related in the north Caucasus Maikop; Masis Blur related in Chalcolithic Armenia; and Mesopotamian Neolithic in the Central Anatolian Bronze Age (Extended Data Fig. 1e,f). These admixtures had begun by around 4300–4000 BC (the date range of the Armenia_C population[13]) and we date them to 4382 ± 63 BC (Extended Data Fig. 2e). The Pre-Pottery Neolithic population of Çayönü was genetically halfway between that of Mardin[14], 200 km to the east, and the Central Anatolian Pottery Neolithic at Çatalhöyük[38] along the Mesopotamian–Anatolian cline. Chalcolithic/Bronze Age people from Southeastern and Central Anatolia all stemmed from the same Çatalhöyük–Mardin continuum, (Supplementary Information section 2). If the proto-Anatolians came from the east, their descendants may have been at the state of Armi, the precise location of which is uncertain but whose Anatolian personal names are recorded by their neighbours in the kingdom of Ebla in Syria[5] in the 25th century BC, half a millennium before Anatolian languages are attested, and just south of the proposed migratory path (Extended Data Fig. 1f). We therefore propose that people of the CLV cline migrated southwards in around 4400 BC, a millennium before the Yamnaya, admixed along the way, and finally reached Central Anatolia from the east.

We find Y-chromosome evidence consistent with this reconstruction: there are sporadic instances of steppe-associated Y-chromosome haplogroup R-V1636 in West Asia at Arslantepe[15] in eastern Anatolia and in Kalavan[13] in Armenia in the Early Bronze Age (around 3300–2500 BC) among individuals without detectible steppe ancestry in the rest of their genomes[10,13]. The R-V1636 individual (ART038) from Arslantepe does not clearly have BPgroup ancestry (3.6 ± 3.1%), but ART027 from the same site (3370–3100 BC) does (16.7 ± 3.5%; $P = 0.171$), preceding the same mix in Early Bronze Age Central Anatolia by a few centuries. R-V1636 in the Remontnoye male, both of those from Progress-2 (ref. 9), two of three from Berezhnovka and 11 individuals of the Volga cline show it to be a prominent lineage of the pre-Yamnaya steppe, and it also appeared as far away as northern Europe[39,40]. A single R-V1636 individual (SA6010; 2886–2671 BC) from Sharakhalsun[9], consistent with CLV ancestry (Fig. 2), is found post-Yamnaya, a last hold-out of this once pervasive lineage (Fig. 3).

The Yamnaya expansion

We infer the average date of mixture in Core Yamnaya[41] to be 4038 ± 48 BC (Extended Data Fig. 2a), with sources related to UNHG/EHG hunter-gatherers and West Asian/Caucasus-related people (Fig. 1b). Such a date does not preclude the possibility that the mixture began earlier or continued afterwards, but it corresponds strikingly to the burgeoning of the Serednii Stih culture. The ancestors of the Core Yamnaya (Fig. 1b and Extended Data Table 2) must have been geographically

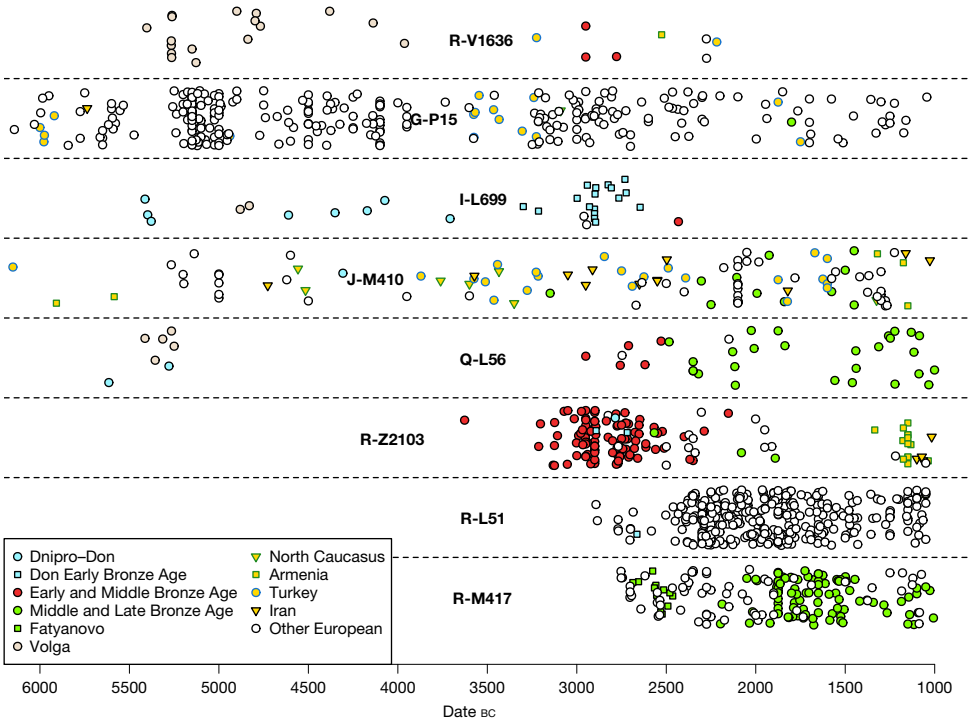


Fig. 3 | Patrilineal succession. Temporal distribution of key Y-chromosome haplogroups from Kazakhstan, Kyrgyzstan, Mongolia, Russia, Turkmenistan, Ukraine, Uzbekistan and comparative regions of Europe and West Asia, 6000–1000 BC. The Early and Middle Bronze Age group includes the Yamnaya, Afanasievo, Poltavka, Catacomb, Chemurchek and North Caucasus cultures; the Middle and Late Bronze Age group includes individuals of diverse cultures down to 1000 BC, including those of the Sintashta, Andronovo, Potapovka and Srubnaya cultures. Information on which individuals are plotted can be found in Supplementary Information (Supplementary Table 6).

constrained[17], contrasting with their later distribution from China to Hungary (Extended Data Fig. 3a, Extended Data Table 2 and Supplementary Table 6), even while maintaining high genetic similarity (mean F_{ST} = 0.005) (Extended Data Table 3). The Don Yamnaya (Extended Data Fig. 3a) are modelled as 79.4 ± 1.1% Core Yamnaya and 20.6 ± 1.1% UNHG. The non-Yamnaya component may be underestimated, if, as is plausible, the Core Yamnaya admixed with a Serednii Stih population of partial UNHG ancestry. We estimate that the Don Yamnaya formed in the late fourth millennium BC (Extended Data Fig. 2b), when, one may assume, unmixed UNHG were rare.

The western expansion also brought Yamnaya into southeastern Europe, reaching as far as Albania and Bulgaria[3,10]. Many of these cluster with the Core Yamnaya, but others deviate towards Neolithic and Chalcolithic populations of southeastern and central Europe (Extended Data Fig. 3b). Yamnaya admixture with these (Extended Data Table 4) occurred in the late fourth millennium BC (Extended Data Fig. 2c), after sporadic early Chalcolithic migrations into southeastern Europe from the steppe[3,25]. By contrast, the Don Yamnaya expanded little, because almost no individuals with high-quality data outside the Don are a clade with them (Supplementary Information section 2); the lower Don was a cul-de-sac for the Yamnaya expansion.

Y-chromosome haplogroup sharing is not informative for Core Yamnaya origins but shows that the Don Yamnaya, dominated by haplogroup I-L699 (17 of 20 instances), had continuity with their Serednii Stih and Neolithic hunter-gatherer ancestors (Fig. 3 and Supplementary Table 7). The Core Yamnaya had R-M269 (49 of 51 instances), most of which was the R-Z2103 (41 of 51) sublineage, which was undetected before the Yamnaya period and related to R-L51, prevalent among Bell Beaker burials[7] and non-steppe Europe (Fig. 3). Slightly more distant is R-PF7563, found in Mycenaean Greece[42]. R-L23, formed at around 4450 BC (https://www.yfull.com/tree/R-L23/; v.12.04.00), unifies in the Eneolithic Beakers, Yamnaya and Mycenaeans. Population divergences are lower than haplogroup ones, so these lineages may have coexisted within the Yamnaya. Finding the R-L23 founder population remains challenging, but our failure to sample it thus far is not surprising if it was small and isolated.

That the Core Yamnaya are part of the Dnipro cline may indicate an origin in the Dnipro basin itself. However, the Dnipro cline is generated by admixture with Dnipro–Don people (UNHG/GK2 related), and the Yamnaya on the Don are also part of this cline, so an alternative origin in the Don area cannot be excluded. Solutions further east are unlikely because the Yamnaya are on neither the Volga nor the CLV cline. The situation is similar for solutions west of the Dnipro: the Core Yamnaya have little or no European farmer ancestry (from the west)[17] (Fig. 1b). A more western origin of the Core Yamnaya would also bring their latest ancestors in proximity with the likely founders of the Corded Ware complex, whose origin is itself in question but who must have been in the area of central eastern Europe occupied by the Globular Amphora culture west of the Core Yamnaya. Most Corded Ware individuals, who can be fit as tracing a large part of their ancestry to the Yamnaya[2,12], were formed by admixture concurrent with the Yamnaya expansion[41] (Extended Data Fig. 2d), shared identical-by-descent (IBD) segments demonstrating genealogical timeframe connections[43], and had a balance of ancestral components for their non-European farmer-related ancestry that was indistinguishable from the Yamnaya[6]. The early-third millennium BC history of the Corded Ware population is intertwined with the Yamnaya expansion because it involved admixture with genetically, if not necessarily archaeologically, Yamnaya people. The Dnipro–Don area of the Serednii Stih culture fits the genetic data, because it explains the ancestry of the nascent Core Yamnaya. All ancestral components found in the Serednii Stih and lacking elsewhere are found in the Yamnaya (Extended Data Fig. 4), and from the Dnipro–Don area, both Corded Ware and southeastern European Yamnaya in the west, and the Don Yamnaya in the east, could have emerged by admixture of the Core Yamnaya with European farmers and UNHG descendants, respectively.

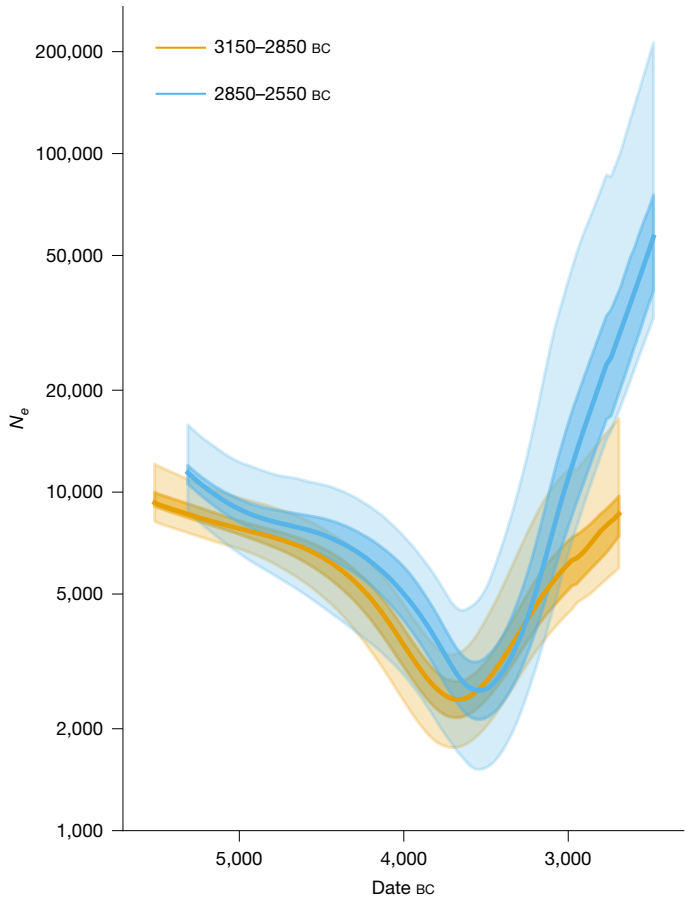


Fig. 4 | Trajectory of the Yamnaya expansion. We used HapNe-LD to estimate the changes in effective population size (N_e) over time of Yamnaya ancestors. We carried out the computation separately for the individuals from the earlier 300 years of our sampling and the later 300 years; shading shows confidence intervals (dark, 50%; light, 95%). Jointly displaying these two trajectories reveals an extraordinary population expansion at 3642–3374 BC (intersection of 95% confidence intervals for the two analyses for the minimum), from when the effective size is a few thousand to an order of magnitude larger. The offset on the *x* axis is due to the difference in sampling time between the two groups.

We estimated the population growth of Core Yamnaya using HapNe-LD, which infers effective population-size fluctuations in low-coverage ancient DNA data[44]. Core Yamnaya dating to the first 300 (n = 25) and later 300 (n = 26) years of our sampling produce 95% confidence intervals of 3829–3374 BC and 3642–3145 BC for the time before growth (Fig. 4). For both, these correspond to growth from an effective number of reproducing individuals of a few thousand. These intervals overlap at 3642–3374 BC, which is the late Serednii Stih period. Taken together with the admixture dating, a scenario emerges in which Yamnaya ancestors were formed by admixture at around 4000 BC, and half a millennium later, a subgroup of them developed or adopted cultural innovations, expanded dramatically and manifested archaeologically around 3300 BC.

IBD[43] genomic segments of at least 20 cM between pairs of individuals did exist before the Yamnaya between regional populations (Fig. 5a), but they became much more common in the Yamnaya period (Fig. 5b). Segments shared across more than 500 km were extremely rare before the Yamnaya (Fig. 5c), but were a few percent between 500 and 5,000 km (Fig. 5d) in the Yamnaya period. Close genetic relatives, sharing at least three segments of at least 20 cM (about fifth-degree relatives)[43] or a sum of IBD of 100 cM or more, were found within 500 km in both periods, and at a much higher rate within each cemetery (Fig. 5e,f). Around 14.4% of Yamnaya–Afanasievo individual pairs within kurgans were close relatives, and 7.4% of them across kurgans of the same cemetery, which is much lower than the 29.0% in the tightly connected pedigree of the Hazleton North chambered tomb in Neolithic Britain from around 3700 BC[45] (P = 0.00075; Fisher's exact test). Kurgans were therefore not family tombs[46] of biological relatives; indeed, biological kinship in them was mostly due to common descent centuries in the past, and close kinship links within kurgans were largely non-biological.

The origin of Indo-Anatolian languages

The conventional view defines Indo-European as including Anatolian languages as the first split[47,48]. Here we use a newer terminology that denotes the entire group as Indo-Anatolian and restricts Indo-European to the related non-Anatolian language families, including Tocharian, Greek and Sanskrit[4,10]. The split of Indo-Anatolian is linguistically dated to 4300–3500 BC[4,37,48,49], pre-dating both the attestation of the Hittite language in Central Anatolia (post-2000 BC[4]) and the Yamnaya expansion. We identify the Yamnaya as the proto-Indo-Europeans for several reasons: first, the formation of the Yamnaya around 4000 BC and their expansion from the end of the fourth millennium BC corresponds to the Indo-European–Anatolian split. Second, the Afanasievo migration[12], plausibly carrying languages ancestral to Tocharian, is widely recognized as the second, post-Anatolian, split[50]. The Yamnaya contributed, after 2500 BC, to Armenians and, since the Early Bronze Age (Extended Data Table 2c), to the Balkans[3,10], where Greek and lesser-known Balkan Indo-European languages such as Illyrian and Thracian were spoken[10,35,42]. For the remaining Indo-European languages, transmission was indirect via descendant cultures of mixed Yamnaya–European farmer origins expanding well beyond the steppes. It is from them that the vast majority of present-day Indo-Europeans are descended. These include non-Balkan European (Italic, Celtic, Germanic, Baltic, and Slavic) speakers through the geographically complementary Corded Ware[2,12] and Beaker cultures of the third millennium BC[7]. The Indo-Iranians, the largest surviving Indo-European group of Asia, were ultimately descended from the Corded Ware too, via a long chain of eastward migrations to Fatyanovo[51] and Sintashta[8,34].

Yamnaya and Anatolians share CLV ancestry (Fig. 2e,f), which must stem from proto-Indo-Anatolian language speakers, except for the possibility of an early transfer of language without admixture. That the CLV ancestry in Central Anatolians during the Hittite presence included lower Volga-related ancestry implies an origin north of the Caucasus (Fig. 2f and Extended Data Fig. 1). Long (30 cM or longer) IBD segments shared by Igren-8 Serednii Stih and Areni-1 with Berezhnovka-2 document Eneolithic links of lower Volga ancestry (Extended Data Table 5), and one link (15.2 cM) between the north Caucasus Vonyucka-1 with early Bronze Age Ovaören (MA2213) ties Central Anatolia to this once expansive network. Even so, only two Indo-Anatolian descendant groups transmitted their languages to posterity: the Yamnaya, aided by their horse-wagon technology[6], and Anatolian speakers, surviving long enough for their languages to be committed to clay around 2000 BC[5], vanishing in late antiquity and fortuitously decyphered in the twentieth century. Our reconstruction, based on genetics (Extended Data Fig. 5), has traced both groups to the CLV people north of the Caucasus, but it cannot discern who first spoke pre-Indo-Anatolian languages.

Linguistic evidence has been advanced in favour of different solutions to the problem of proto-Indo-European origins for more than two centuries, and here we review some recent proposals that are relevant to our reconstruction of early Indo-Anatolian/Indo-European history.

First, cereal terminology in Indo-Anatolian/Indo-European languages may restrict Indo-Anatolian origins to the easternmost extent of agricultural subsistence during the Eneolithic, the Dnipro valley[52]. Our findings do not contradict this, but they raise the possibility of a

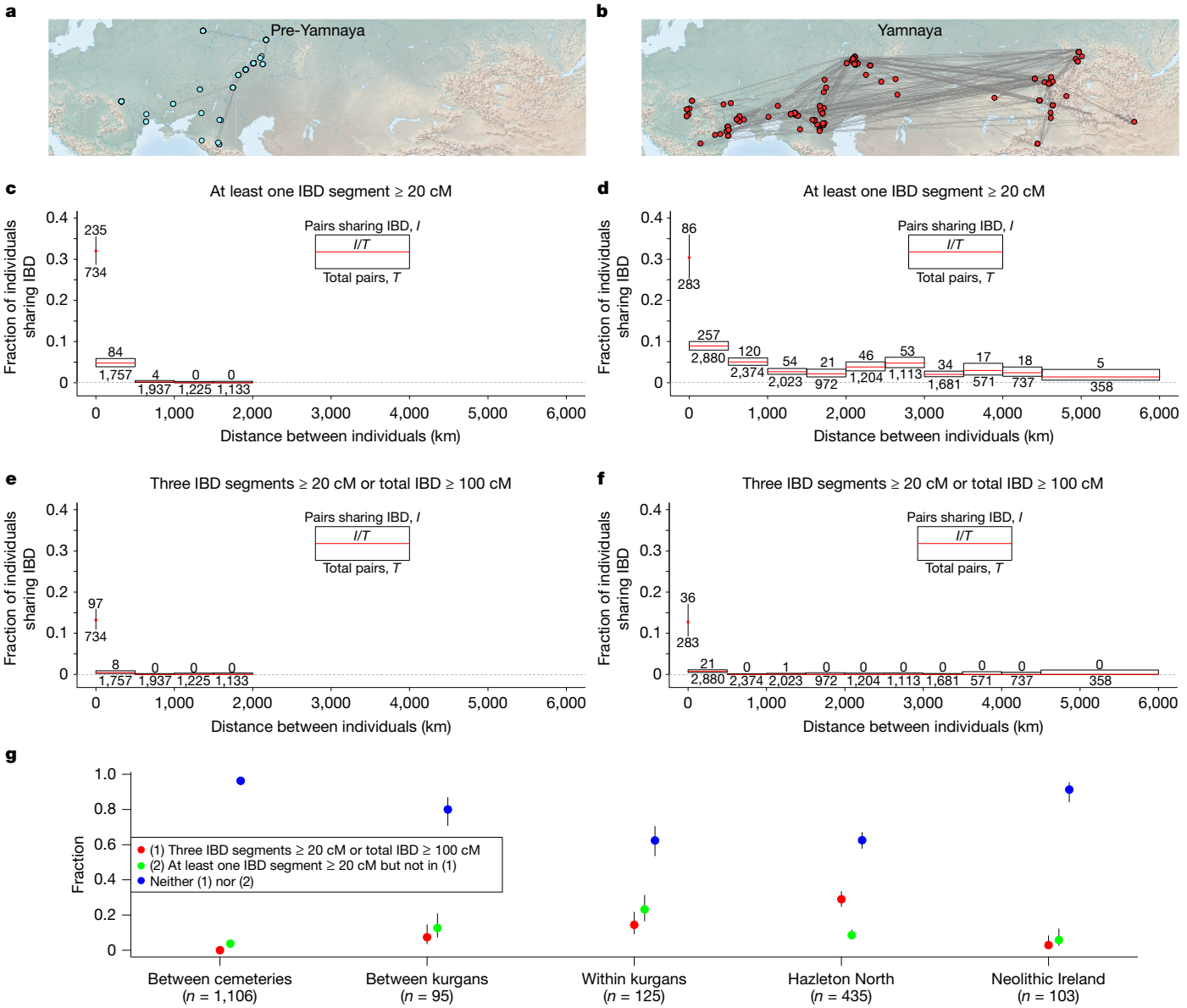


Fig. 5 | IBD analysis of the Yamnaya and their predecessors. a,**b**, Pairs of individuals linked by at least one IBD segment at least 20 cM in length reveal a sparse but highly connected network in the pre-Yamnaya (Methods) (**a**) and Yamnaya (**b**) groups. No detectable IBD is found in the pre-Yamnaya period beyond the scale of 1,000 km. **c**,**d**, Yamnaya share more IBD with each other at short distance scales compared with the pre-Yamnaya people (**c**), but IBD sharing extends all the way to the roughly 6,000 km scale of their geographical distribution (**d**). **e**,**f**, However, closely related individuals occur only at short distance scales in both pre-Yamnaya (**e**) and Yamnaya (**f**) groups, indicating that IBD sharing in the Yamnaya was a legacy of their common origin. In **c**–**f**, two-sided 95% confidence intervals are shown as a vertical interval (at distance = 0) or a rectangle (at distance ranges greater than 0); the fraction of number of pairs of individuals sharing IBD (*I*)/total number of pairs of individuals (*T*) is shown in red. **g**, In a set of 9 Yamnaya cemeteries and a total of 25 kurgans, closely or distantly related individuals are almost absent in inter-cemetery comparisons, more are found in inter-kurgan and within-cemetery comparisons, and even more are found in intra-kurgan comparisons; nonetheless, most Yamnaya individuals in all comparisons were unrelated. Kurgan burial of close kin was less common than in the case of a local patrilineal dynasty, as at a Neolithic long cairn at Neolithic Hazleton North[45], but was more common than in Neolithic monuments in Ireland[55]. Two-sided 95% confidence intervals are shown. The map was drawn using public-domain Natural Earth data with the rnaturalearth package in R[54].

Caucasus (rather than European) Neolithic source for this vocabulary through the CLV cline.

Second, the attestation of Anatolian languages largely in central-western Anatolia can be explained most parsimoniously by a western entry (through the Balkans)[4], but genetic data provide strong evidence in favour of an eastern route[53], because not only CLV but especially Mesopotamian Neolithic, the two sources of the Central Anatolian Bronze Age groups, are eastern. Further evidence comes from observing no European farmer or hunter-gatherer ancestry in Central Anatolian Bronze Age groups, as might be expected from a Balkan route from the west, although if these groups by-passed local Europeans, or used a maritime route, we would not see European mixture. A weakness of the eastern-entry hypothesis has always been that there is no linguistic evidence of Anatolian speakers in eastern Anatolia along the proposed migratory path. However, this argument does not add relative weight to the western-entry hypothesis either, because no linguistic evidence for migratory pre-Anatolian speakers has been found in the southeastern European path proposed by that hypothesis. The lack of linguistic traces in eastern Anatolia could be explained by the archaeologically momentous expansion of the Kura-Araxes archaeological

culture in the Caucasus and eastern Anatolia after around 3000 BC, which may have driven a wedge between steppe and West Asian speakers of Indo-Anatolian languages, isolating them from each other and perhaps explaining their survival in western Anatolia into recorded history. That the expansion of the Kura-Araxes archaeological culture could have had a profound enough demographic impact to have pushed out Anatolian speakers is directly attested by genetic evidence showing that, in Armenia, the spread of the Kura-Araxes culture was accompanied by the complete disappearance of CLV ancestry that had appeared there in the Chalcolithic[10,13] (Fig. 2f).

The Kura-Araxes culture may not be the only reason for the Indo-Anatolian split. Autosomal and Y-chromosome homogenization of the Yamnaya ancestral population in the fourth millennium BC provides another lens through which to understand its origins, with isolation fostering linguistic divergence. This may have persisted after its expansion: previous inhabitants largely disappear in the face of the Yamnaya juggernaut, albeit with exceptions[17]. Perhaps mixing, which was avoided by the kurgan elites, occurred between locals and Yamnaya not buried in kurgans. The rise of the Yamnaya on the steppe at the expense of their predecessors was followed by their demise after about 1,000 years, displaced by descendants of people of the Corded Ware culture. Was this the fall of the kurgan elites or of the population as a whole? The steppe was dominated by many and diverse groups later still, such as the Scythians and Sarmatian nomads of the Iron Age. These groups were certainly diverse genetically, but their kurgans, found across the steppe, attest to the persistence of at least some elements of culture that began in the Caucasus–Volga area some 7,000 years ago before blooming, in the Dnipro–Don area, into the Yamnaya culture that first united the steppe and had an impact on most of Eurasia. For what symbolic purpose the Yamnaya and their precursors erected these mounds we may never fully know. If they aimed to preserve the memory of those buried under them, they did achieve their goal, as the kurgans, dotting the landscape of the Eurasian steppe, drew generations of archaeologists and anthropologists to their study, enabling the genetic reconstruction of their makers' origins presented here.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08531-5.

1. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
2. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
3. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
4. Kloekhorst, A. in *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics* (eds Kristiansen, K. et al.) 42–60 (Cambridge Univ. Press, 2023).
5. Kroonen, G., Barjamovic, G. & Peyrot, M. Linguistic supplement to Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian. *Zenodo* https://doi.org/10.5281/zenodo.1240524 (2018).
6. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2007).
7. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
8. Narasimhan V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
9. Wang, C.-C. et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
10. Lazaridis, I. et al. The genetic history of the Southern Arc: a bridge between West Asia and Europe. *Science* **377**, eabm4247 (2022).
11. Allentoft, M. E. et al. Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311 (2024).
12. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
13. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
14. Lazaridis, I. et al. Ancient DNA from Mesopotamia suggests distinct Pre-Pottery and Pottery Neolithic migrations into Anatolia. *Science* **377**, 982–987 (2022).
15. Skourtanioti, E. et al. Genomic history of Neolithic to Bronze Age Anatolia, Northern Levant and Southern Caucasus. *Cell* **181**, 1158–1175 (2020).
16. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
17. Nikitin, A. G. et al. A genomic history of the North Pontic Region from the Neolithic to the Bronze Age. https://doi.org/10.1038/s41586-024-08372-2 (2025).
18. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
19. Zeng, T. C. et al. Postglacial genomes from foragers across Northern Eurasia reveal prehistoric mobility associated with the spread of the Uralic and Yeniseian languages. *Nature* (in press).
20. Posth, C. et al. Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature* **615**, 117–126 (2023).
21. Vybornov, A. et al. Diet and chronology of Neolithic–Eneolithic cultures (from 6500 to 4700 cal BC) in the Lower Volga Basin. *Radiocarbon* **60**, 1597–1610 (2018).
22. Gimbutas, M. *The Prehistory of Eastern Europe* (Peabody Museum, 1956).
23. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
24. Anthony, D. W. et al. The Eneolithic cemetery at Khvalynsk on the Volga River. *Praehist. Z.* **97**, 22–67 (2022).
25. Penske, S. et al. Early contact between late farming and pastoralist societies in southeastern Europe. *Nature* **620**, 358–365 (2023).
26. Nikitin, A. G., Ivanova, S., Culleton, B. J., Potekhina, I. & Reich, D. New radiocarbon and stable isotope data from the Usatove culture site of Mayaky in Ukraine. *SSRN* https://doi.org/10.2139/ssrn.4236123 (2023).
27. Govedarica, B. & Manzura, I. The Giurgiulesti cemetery in chronological and cultural context of Southeastern and Eastern Europe. *Eurasia Antiqua* **22**, 1–39 (2016).
28. Skorobogatov, A. M. Pamyatniki Neolita I Eneolita v Ust'e Chernoi Kalitvy. *Trudy Voronezhskogo Oblastnogo Kraevedcheskogo Muzeiya Vyp.* **3**, 47–53 (2019).
29. Skorobogatov, A. M. & Smol'janinov, R. V. Srednestogovskie materialy v bassejne Verhnego i Srednego Dona. *Rossiyskaya arkheologiya* **2013**, 126–136 (2013).
30. Shishlina, N. I. et al. Paleoecology, subsistence, and ^{14}C chronology of the Eurasian Caspian steppe Bronze Age. *Radiocarbon* **51**, 481–499 (2009).
31. Korenevskii, S. N. *Rozhdenie Kurgana* (Taus, 2012).
32. Zhur, K. V. et al. Human DNA from the oldest Eneolithic cemetery in Nalchik points the spread of farming from the Caucasus to the Eastern European steppes. *iScience* https://doi.org/10.1016/j.isci.2024.110963 (2024).
33. Altınışık, N. E. et al. A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic transition. *Sci. Adv.* https://doi.org/10.1126/sciadv.abo3609 (2022).
34. de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
35. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
36. Lazaridis, I. The evolutionary history of human populations in Europe. *Curr. Opin. Genet. Dev.* **53**, 21–27 (2018).
37. Kassian, A. S. et al. Rapid radiation of the inner Indo-European languages: an advanced approach to Indo-European lexicostatistics. *Linguistics* **59**, 949–979 (2021).
38. Yaka, R. et al. Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Curr. Biol.* https://doi.org/10.1016/j.cub.2021.03.050 (2021).
39. Egfjord, A. F.-H. et al. Genomic steppe ancestry in skeletons from the Neolithic single grave culture in Denmark. *PLoS ONE* **16**, e0244872 (2021).
40. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
41. Chintalapati, M., Patterson, N. & Moorjani, P. The spatiotemporal patterns of major human admixture events during the European Holocene. *eLife* **11**, e77625 (2022).
42. A genetic probe into the ancient and medieval history of Southern Europe and West Asia. *Science* **377**, 940–951 (2022).
43. Ringbauer, H. et al. Accurate detection of identity-by-descent segments in human ancient DNA. *Nat. Genet.* **56**, 143–151 (2024).
44. Fournier, R., Tsangalidou, Z., Reich, D. & Palamara, P. F. Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nat. Commun.* **14**, 7945 (2023).
45. Fowler, C. et al. A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* **601**, 584–587 (2022).
46. Shishlina, N. *Reconstruction of the Bronze Age of the Caspian steppes: Life Styles and Life Ways of Pastoral Nomads* (British Archaeological Reports, 2008).
47. Olander, T. Indo-European cladistic nomenclature. *Indogermanische Forschungen* **124**, 231–244 (2019).
48. Anthony, D. W. & Ringe, D. The Indo-European homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist.* **1**, 199–219 (2015).
49. Pronk, T. C. Indo-European secondary products terminology and the dating of Proto-Indo-Anatolian. *J. Indo-European Stud.* **49**, 141–170 (2022).
50. Ringe, D., Warnow, T. & Taylor, A. Indo-European and computational cladistics. *Trans. Philol. Soc.* **100**, 59–129 (2002).
51. Saag, L. et al. Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. *Sci. Adv.* **7**, eabd6535 (2021).
52. Kroonen, G., Jakob, A., Palmér, A. I., van Sluis, P. & Wigman, A. Indo-European cereal terminology suggests a Northwest Pontic homeland for the core Indo-European languages. *PLoS ONE* **17**, e0275744 (2022).
53. Kristiansen, K. in *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European* (eds Serangeli, M. & Olander, T.) 157–165 (Brill, 2019).
54. Massicotte, P. & South, A. rnaturalearth: World map data from Natural Earth. https://docs.ropensci.org/rnaturalearth/ (2024).
55. Cassidy, L. M. et al. A dynastic elite in monumental Neolithic society. *Nature* **582**, 384–388 (2020).

[1]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]Department of Genetics, Harvard Medical School, Boston, MA, USA. [3]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [4]Hartwick College, Department of Anthropology, Oneonta, NY, USA. [5]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czechia. [6]Department of Statistics, University of Oxford, Oxford, UK. [7]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [8]BIOMICs Research Group, Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain. [9]Ikerbasque-Basque Foundation of Science, Bilbao, Spain. [10]Samara State University of Social Sciences and Education, Samara, Russia. [11]Center of Human Ecology, Institute of Ethnology and Anthropology, Russian Academy of Science, Moscow, Russia. [12]Department of Archaeology, State History Museum, Moscow, Russia. [13]"Gavrilă Simion" Eco-Museum Research Institute, Tulcea, Romania. [14]Historical Ecological and Cultural Association Povolzje, Samara Regional Public Organization, Samara, Russia. [15]Azov History, Archaeology and Palaeontology Museum-Reserve, Azov, Russia. [16]Institute of Archaeology named after A.Kh Margulan, Almaty, Kazakhstan. [17]Department of Biological Anthropology, Institute of Biology, University of Szeged, Szeged, Hungary. [18]Research Institute and Museum of Anthropology, Moscow, Russia. [19]Institute of Archeology named after A. Kh. Khalikov Tatarstan Academy of Sciences, Kazan, Russia. [20]Orheiul Vechi Cultural-Natural Reserve, Institute of Bioarchaeological and Ethnocultural Research, Chișinău, Republic of Moldova. [21]Fr. I Rainer Institute of Anthropology, University of Bucharest, Bucharest, Romania. [22]Damjanich János Museum, Szolnok, Hungary. [23]Department of Archaeology, University of Szeged, Szeged, Hungary. [24]Déri Museum, Debrecen, Hungary. [25]Department of Regional Studies of Russia, National and State-Confessional Relations, Altai State University, Barnaul, Russia. [26]Department of Anthropology, Hungarian Natural History Museum-Hungarian National Museum Public Collection Centre, Budapest, Hungary. [27]Research Institute GAUK RO "Don Heritage", Rostov-on-Don, Russia. [28]Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, České Budějovice, Czechia. [29]Prahova County Museum of History and Archaeology, Ploiești, Romania. [30]Department of Geography, Faculty of Humanities, University Valahia of Târgoviște, Târgoviște, Romania. [31]Department of Biological Anthropology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary. [32]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria. [33]Human Evolution and Archaeological Sciences, University of Vienna, Vienna, Austria. [34]Department of Mountain and Highland Archaeology, Institute of Archaeology and Ethnology, Polish Academy of Science, Kraków, Poland. [35]Slovak National Museum-Archaeological Museum, Bratislava, Slovak Republic. [36]Peter the Great Museum of Anthropology and Ethnography, Department of Physical Anthropology, St. Petersburg, Russia. [37]State Autonomous Cultural Institution of the Republic of Khakassia "Khakassian National Museum of Local Lore named after L.R. Kyzlasova", Abakan, Russia. [38]Institute of Archaeology, HUN-REN Research Centre for the Humanities, Budapest, Hungary. [39]Centre for Egyptological Studies of the Russian Academy of Sciences, Russian Academy of Sciences, Moscow, Russia. [40]Department of Archaeology and History of the Ancient World, Southern Federal University, Rostov-on-Don, Russia. [41]Museum of Vojvodina, Novi Sad, Serbia. [42]Department of History of the Institute of Humanities, Ural Federal University, Ekaterinburg, Russia. [43]Institute of Plant and Animal Ecology, Urals Branch of the Russian Academy of Sciences, Yekaterinburg, Russia. [44]Institute of History, Archaeology and Ethnography, Dagestan Branch of the Russian Academy of Science, Makhachkala, Dagestan, Russia. [45]Independent researcher, Philadelphia, PA, USA. [46]Department of General History, Historical and Literary Institute of the State University of Education, Ministry of Education Moscow, Moscow, Russia. [47]Centre for Applied Bioanthropology, Institute for Anthropological Research, Zagreb, Croatia. [48]Department of Archaeology and Heritage, Faculty of Humanities, University of Primorska, Koper, Slovenia. [49]Kalmyk Scientific Centre of the Russian Academy of Sciences, Elista, Russia. [50]National Agency for Archaeology, Chișinău, Republic of Moldova. [51]National Research Tomsk State University, Tomsk, Russia. [52]V.F. Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, Russia. [53]Department of Archaeology, Ethnography and Museology, Altai State University, Barnaul, Russia. [54]Laboratory of Ancient and Medieval Archaeology of Eurasia, Altai State University, Barnaul, Russia. [55]Slovak National Museum-Natural History Museum, Bratislava, Slovak Republic. [56]Faculty of History, University of Oxford, Oxford, UK. [57]Center for Stone Age Archeology, Institute of History and Archaeology, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia. [58]Olga Necrasov Centre for Anthropological Research, Romanian Academy, Iași Branch, Iași, Romania. [59]Tyumen Scientific Center of the Siberian Branch of Russian Academy of Sciences, Institute of Problems of Northern Development, Tyumen, Russia. [60]Institute for the History of Material Culture, Russian Academy of Sciences, St Petersburg, Russia. [61]Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow, Russia. [62]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. [63]Museo delle Civiltà, Italian Ministry of Culture, Rome, Italy. [64]School of Archaeology, University College Dublin, Dublin, Ireland. [65]Institute of Archaeogenomics, HUN-REN Research Centre for the Humanities, Budapest, Hungary. [66]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [67]These authors contributed equally: Iosif Lazaridis, Nick Patterson, David Anthony, Leonid Vyazov. ✉e-mail: lazaridis@fas.harvard.edu; nickp@broadinstitute.org; AnthonyD@hartwick.edu; l.a.vyazov@gmail.com; ron.pinhasi@univie.ac.at; reich@genetics.med.harvard.edu

Methods

Terminology for archaeological cultures and geographic locations

For archaeological cultures and geographic locations that span more than one modern country, we used the prevalent term in the archaeological and genetic literature; for example, Yamnaya is the common term in Russia and most of Eastern Europe, instead of the Ukrainian Yamna. For archaeological cultures and locations that are confined to a single country, we generally used the local terminology; for example, we referred to the archaeological cultures of Usatove, Trypillia and Serednii Stih, and the river Dnipro, using the Ukrainian terms instead of the corresponding Russian terms Usatovo, Tripolye, Sredni Stog and Dniepr.

Previously published Eneolithic and Yamnaya/Afanasievo individuals

We counted previously published Yamnaya/Afanasievo individuals with genome-wide autosomal data (n = 75) from the archaeogenetic literature[2,3,8–10,12,33,56–62]. We counted pre-Yamnaya Eneolithic individuals[3,9,11,17,20,40,51,63,64] with genome-wide data from the European steppe and its environs (n = 82) by filtering individuals to the date range 5000–3500 BC, the countries of Russia and Ukraine, and latitude west or equal to 60° E and longitude south or equal to 60° N.

Sampling ancient individuals

The skeletal remains were all analysed with permission from local authorities in each location from which they came. Every sample is represented by stewards, such as archaeologists or museum curators, who are either authors of this paper or are thanked in the Acknowledgements. The remains were nearly all sampled in ancient-DNA clean rooms, either at Harvard Medical School, the University of Vienna or the Institute for Archaeogenomics in Budapest. We prioritized sampling petrous bones if they were available and accessible, taking bone powder from the cochlea by sandblasting and milling[65], or drilling into the cochlea directly after physical surface cleaning, or drilling through the cranial base to minimize damage to intact skulls[66]. If we could not sample from the cochlea, we sought to sample a tooth, prioritizing the cementum layer after physical surface cleaning[67]. If neither a cochlea nor a tooth was available, we sought to sample a dense cortical bone, which we analysed by drilling and collecting powder after physical surface cleaning. For some samples that could not leave the museum, we sampled on site, either drilling directly into the cochlea, the tooth root or bone after physical surface removal. We sometimes dislodged auditory ossicles during sandblasting or drilling into the cochlea. When this happened during the cleaning procedure, we generally stopped the destructive sampling and collected the ossicle(s)[68]. As suggested in the study[68] that recognized the high preservation of DNA in ossicles, we cleaned the ossicle with 10% bleach and radiated it with UV light for 10 min before submerging it in extraction buffer without attempting to produce powder.

Ancient-DNA data generation

The samples for which we report new data were processed between 2013 and 2023, and so were analysed using a changing set of protocols. Details and protocols used for each library can be found in Supplementary Table 2. At Harvard Medical School, where most of the wet laboratory work was done, we initially carried out all DNA extractions and Illumina library preparations manually, using small batches of samples and silica columns for DNA clean-up[69–71]. Beginning in 2018, we used automated liquid handlers (Agilent Bravo Workstations) for both DNA extraction[72] and library preparation with magnetic beads (see the supplementary material in ref. 73 for automated double-stranded library preparation, and ref. 74 for automated single-stranded library preparation). We treated DNA extracts with USER (NEB) during library preparation to cut DNA at uracils; this treatment is inefficient at terminal uracils and leaves a damage pattern expected for ancient DNA at the terminal bases that can be filtered out for downstream analysis while allowing a library to be authenticated as old. All libraries were either dual barcoded through double-stranded ligation or dual indexed through indexing PCR at the end of single-stranded library preparation to allow pooling before sequencing.

Before 2015, we screened libraries for mitochondrial DNA (mtDNA) before attempting to capture nuclear loci[75]. In the following two years, we added an increasing number (between 10 and 4,000) of nuclear single-nucleotide polymorphisms (SNPs) as targets for the screening capture because mtDNA quality does not always correlate well with nuclear DNA quality and quantity. We later increased the number of targeted SNPs in our nuclear capture from about 390,000 (390k)[2,76] to about 1.24 million (1,240k)[77] for libraries passing the mitochondrial capture with nuclear spike-in. We later stopped using the screening capture and added the mitochondrial probes to the 1240k probes (1240k+). In 2022, we switched from the 1240k homebrew capture to a kitted capture product available from Twist Biosciences[78].

For ancient-DNA data generated at the Institute of Archaeogenomics in Budapest, we followed the protocol described in ref. 79.

Bioinformatic processing

All ancient-DNA libraries were sequenced with paired-end reads on Illumina instruments. We then performed the following steps: preprocessing, alignment and post-alignment filtering for variant calling. The goal of preprocessing was to take raw sequenced products and create merged sequences for alignment. We demultiplexed reads, binned these to whichever library each read belonged to using the identifying barcodes and indices, trimmed these identifying markers as well as any residual adapter sequences, and merged each paired-end read into a single molecule using the overlap of the paired-end reads as a guide, employing a modified version of SeqPrep (https://github.com/jstjohn/SeqPrep). We aligned the resulting sequences to both the hg19 human genome reference (https://www.internationalgenome.org/category/grch37/) and the inferred ancestral Reconstructed Sapiens Reference Sequence mitochondrial sequence[80] using the samse aligner of bwa[81]. We marked duplicate molecules by barcode bin, based on the same start and stop positions and orientation. The computational pipelines with specific parameters used are available on GitHub at https://github.com/dReichLab/ADNA-Tools and https://github.com/dReichLab/adna-workflow.

We used a pseudohaploid genotyping approach to determine a randomly selected allele at SNP sets of interest. To represent the allele at each SNP, we randomly selected sequences from a pool of all sequences covering that position with a minimum data quality; our criteria were a minimum mapping quality of at least 10 and a base quality of at least 20, after trimming sequences by 2 base pairs at both the 5′ and 3′ ends to remove damage artefacts. We assessed ancient-DNA authenticity by using contamMix-1.0.1051[82] to search for heterogeneity in mtDNA sequences, which are expected to be non-variable in uncontaminated individuals, and also ANGSD to test for heterogeneity in X-chromosome sequences, which are expected to be homozygous in males[83]. We further evaluated the authenticity of the ancient samples by using pmdtools[84] to measure the rate of cytosine-to-thymine mutations in the first and last nucleotides (in untrimmed sequences), which is expected for genuine ancient DNA[70], and by computing the ratio of Y chromosomes to the sum of X and Y chromosome sequences, which is expected to be very low for females and to have a much higher value for males. We determined a consensus for mtDNA using bcftools (https://github.com/samtools/bcftools) and SAMTools[85], requiring a minimum of two-fold coverage to call the nucleotide and a majority rule to determine its value. We used HaploGrep2 to determine mitochondrial haplogroups based on the phylotree database (mtDNA tree build 17)[86,87].

PCA

We projected individuals in Fig. 1b in smartpca[88] using parameters newshrink: YES and lsqporject: YES on a PCA space with axes formed by the following populations: OberkasselCluster (a set of trans-Alpine WHG individuals identified in ref. 20), Russia_Firsovo_N, Iran_HajjiFiruz_C[8], Iran_C_SehGabi[13], Iran_C_TepeHissar[89], Israel_C[90] and Germany_EN_LBK[2,40,79,91]. The coordinates of plotted points are shown in Supplementary Table 5.

F_{ST} estimation

We computed F_{ST} in smartpca[88] with parameters inbreed: YES and fstonly: YES[92].

Drawing maps

We drew the maps in Figs. 1 and 5, Extended Data Figs. 1 and 5, and Supplementary Information section 2 using public-domain Natural Earth data with the rnaturalearth package in R[54]. Digital elevation maps in Supplementary Information section 1 were drawn using the Copernicus digital elevation model (https://doi.org/10.5270/ESA-c5d3d65).

Visualizing the three Eneolithic clines and preceding populations

We fit models for Eneolithic cline populations (Fig. 1c) using qpAdm[2] and with the following set of right populations: OldAfrica, Russia_AfontovaGora3, CHG, Iran_GanjDareh_N, Italy_Villabruna, Russia_Sidelkino.SG and Turkey_N (Fig. 1c). Diverse ternary models of preceding, Eneolithic and Bronze Age populations are shown in Fig. 2. Individuals plotted at the triangle edge fit the simpler two-source model ($P > 0.05$) (in some of these cases, the three-source models have a negative coefficient from one of the three sources). The corners of each triangle represent the sources. Unplotted individuals all gave fits at $P < 0.05$ and so should be viewed as poorly described by the model.

Model competition with qpAdm/qpWave

We used qpWave/qpAdm methods[2,18] to characterize relationships among diverse target and source populations from the steppe and adjacent areas (Supplementary Information section 2). We use OldAfrica, Russia_AfontovaGora3, CHG, Iran_GanjDareh_N, Italy_Villabruna, Russia_Sidelkino.SG and Turkey_N as the set of right populations for most analyses. For analysis of Anatolians, we expanded this to OldAfrica, CHG, Iran_GanjDareh_N, Italy_Villabruna, Russia_AfontovaGora3, Russia_Sidelkino.SG, TUR_Marmara_Barcın_N, TUR_C_Boncuklu_PPN and TUR_C_Çatalhöyük_N, Natufian to gain leverage for differentiating among West Asian sources. For faster computation, we ran qpWave/qpAdm on precomputed output from qpfstats runs (https://github.com/DReichLab/AdmixTools/blob/master/qpfs.pdf) with a poplistname that includes Han.DG, and all target, source and right populations, and parameters allsnps: YES, inbreed: NO. We performed separate qpWave/qpAdm runs directly on genotype files as needed when the target or source populations were not present in the qpfstats output with parameter basepop: Han.DG. We identified feasible models as having $P > 0.05$, all standard errors ≤0.1, and admixture proportions ≤2 standard errors from 0 and 1. We removed target or source populations from the right set. Competition of models A and B involves two qpWave/qpAdm runs in which all sources of A\B and B\A (\ denotes set difference) are placed on the right set. Details of all analyses can be found in Supplementary Information section 2.

Y-chromosome haplogroup inference

We used the methodology described in ref. 6, which used the YFull YTree v.8.09 phylogeny (https://github.com/YFullTeam/YTree/blob/master/ytree/tree_8.09.0.json) to denote Y-chromosome haplogroups in terminal notation[93].

Estimates of dates of admixture

We used DATES[8,41] to estimate dates of admixture for the Core Yamnaya, Don Yamnaya, Eastern European Yamnaya, Corded Ware and Caucasus–Anatolian populations (Extended Data Fig. 2). For the Core Yamnaya and Caucasus–Anatolian populations, we used sets of diverse West Asian and European hunter-gatherer populations as the two sources. For the Don Yamnaya, we used the Core Yamnaya and UNHG as the two sources. For the Eastern European Yamnaya, we used the Core Yamnaya and a diverse set of Neolithic/Chalcolithic European farmers from Extended Data Fig. 3b. For the Corded Ware, we used the Core Yamnaya and Globular Amphora as the two sources. It is more important to use many source samples even if they are genetically somewhat drifted to the true ones; picking the wrong sources does not bias the date estimate[41].

IBD segment detection

We used ancIBD[43] to detect IBD segments of length greater than or equal to 8 cM. The pre-Yamnaya individuals plotted in Fig. 5 are from the period 5500–3500 BC.

Estimates of geographical distance

To study the decay of IBD with geographical distance, we estimated the distance between sites on the basis of their latitude and longitude, given in Supplementary Table 4, using the Haversine distance as implemented in distHaversine[94] of the package geosphere in R.

Estimates of effective population size

We ran HapNe-LD (v.1.20230726)[18] using the default parameters and providing pseudo-haploid genotypes as input. In brief, HapNe-LD uses a summary statistic that measures long-range correlations between markers to infer fluctuations in effective population size (defined as the inverse of the coalescence rate) over time. We studied two distinct sets of unrelated individuals, all of which had a coverage of at least 0.7× on the target autosomal SNPs and with a standard deviation on their estimated date smaller than 180 years (about 6 generations). The first group consisted of 25 Core Yamnaya individuals with estimated dates ranging between 4,500 and 4,800 years before present. The second group contained 26 Core Yamnaya individuals ranging from 4,800 to 5,100 years before present.

If no evidence of effective population-size fluctuations can be detected in the data, HapNe-LD produces a flat line. An output containing fluctuations should thus be interpreted as the detection of changes in historical effective population size. Recent admixture between highly differentiated populations ($F_{ST} > 0.1$) might lead to biases in LD-based analyses that induce fluctuations similar to a population bottleneck. However, HapNe implements a test to flag the presence of recent structure in the data, which was not detected in either sample set (approximate $P \geq 0.1$), indicating that the observed signal instead reflected variation in the effective population size of these groups.

In our analyses, the effective population size was defined as the inverse of the instantaneous coalescence rate. This quantity corresponds to twice the number of breeding individuals in an idealized population. As well as changes in the number of individuals in the population (census size), several factors, such as changes in population structure, selection and cultural practices[95], can have an influence on the effective population size. These factors may in part be responsible for the effective size fluctuations observed in the Core Yamnaya.

We inferred approximate confidence intervals using bootstrap with different chromosome arms as resampling units. We determined the beginning of the expansion by using the location of the minimum of each bootstrapped trajectory. We converted the results into years by assuming 28.6 years per generation for the median minimum location, and 25.6 and 31.5 years per generation for the lower and upper bounds, respectively[96]. We used these values, corresponding to the estimated

number of years per generation for males (31.5) and females (25.6), to account for uncertainty in the conversion factor.

Ethics statement. The individuals studied in this work were all analysed with the goal of minimizing damage to their skeletal remains, with permission from local authorities in each location from which they came. Every sample is represented by stewards, such as archaeologists or museum curators, who are either authors or are thanked in the Acknowledgements. Open-science principles require making all the data used to support the conclusions of a study maximally available, and we support these principles here by making publicly available not only the digital copies of molecules (the uploaded sequences), but also the molecular copies (the ancient-DNA libraries, which constitute molecular data storage). Researchers who wish to carry out deeper sequencing of the libraries published in this study can make a request to corresponding author D.R. We commit to granting reasonable requests as long as the libraries remain preserved in our laboratories, with no requirement that we be included as collaborators or co-authors on any resulting publications.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Genotype data for individuals included in this study can be obtained from the Harvard Dataverse repository at https://doi.org/10.7910/DVN/QGNMRH. The DNA sequences reported in this paper have been deposited in the European Nucleotide Archive under accession number PRJEB81467. Other newly reported data, such as radiocarbon dates and archaeological context information, are included in this paper and the Supplementary Information.

56. Järve, M. et al. Shifts in the genetic landscape of the Western Eurasian steppe associated with the beginning and end of the Scythian Dominance. *Curr. Biol.* **29**, 2430–2441 (2019).
57. Jeong, C. et al. A dynamic 6,000-year genetic history of Eurasia's eastern steppe. *Cell* **183**, 890–904 (2020).
58. Kumar, V. et al. Bronze and Iron Age population movements underlie Xinjiang population history. *Science* **376**, 62–69 (2022).
59. Patterson, N. et al. Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* **601**, 588–594 (2022).
60. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
61. Zhang, F. et al. The genomic origins of the Bronze Age Tarim Basin mummies. *Nature* **599**, 256–261 (2021).
62. Wang, C.-C. et al. Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021).
63. Gelabert, P. et al. Genomes from Verteba cave suggest diversity within the Trypillians in Ukraine. *Sci. Rep.* **12**, 7242 (2022).
64. Mattila, T. M. et al. Genetic continuity, isolation, and gene flow in Stone Age Central and Eastern Europe. *Commun. Biol.* **6**, 793 (2023).
65. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* **14**, 1194–1205 (2019).
66. Sirak, K. A. et al. A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques* **62**, 283–289 (2017).
67. Damgaard, P. B. et al. Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep.* **5**, 11184 (2015).
68. Sirak, K. et al. Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Res.* **30**, 427–436 (2020).
69. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
70. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
71. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
72. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
73. Prendergast, M. E. et al. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science* **365**, eaaw6275 (2019).
74. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
75. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).
76. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
77. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
78. Rohland, N. et al. Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res.* **32**, 2068–2078 (2022).
79. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
80. Behar, D. M. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
82. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
83. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
84. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
85. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
86. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
87. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
88. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
89. Shinde, V. et al. An ancient Harappan genome lacks ancestry from steppe pastoralists or Iranian farmers. *Cell* **179**, 729–735 (2019).
90. Harney, É. et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* **9**, 3336 (2018).
91. Rivollat, M. et al. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci. Adv.* **6**, eaaz5344 (2020).
92. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
93. Adamov, D., Gurianov, V. M., Karzhavin, S., Tagankin, V. & Urasin, V. Defining a new rate constant for Y-chromosome SNPs based on full sequencing data. *Russian J. Genet. Geneal.* **7**, 1920–2997 (2015).
94. Sinnott, R. W. Virtues of the Haversine. *Sky Telescope* **68**, 158 (1984).
95. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
96. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).

Acknowledgements We thank A. G. Nikitin for advice and feedback; N. Adamski, R. Bernardos, N. Broomandkhoshbacht, D. Fernandes, M. Ferry, E. Harney, K. Mandl, S. Nordenfelt, K. Stewardson, B. G. Mende and Z. Zhang for laboratory or bioinformatics work; L. Bembeeva, B. Preda-Bălănică, I. Ecsedy, A. I. Gotlib, V. M. Heyd, S. A. Mikhailovich, N. Morgunova, A. Soficaru, S. S. Tur and P. Włodarczak for anthropological work and comments; I. Vyazov for help in adapting illustrations; and A. A. Shalapinin, A. F. Kochkina and D. A. Stashenkov for archaeological fieldwork and sample collection. T.Ha. was supported by the Hungarian Research, Development and Innovation Office (grant FK128013), the Bolyai Scholarship of the Hungarian Academy of Sciences, and by the ÚNKP-23-5 New National Excellence Program of the Ministry for Culture and Innovation from the National Research, Development and Innovation Fund. P.F. and L.V. were supported by the Czech Ministry of Education, Youth and Sports (programme ERC CZ, project LL2103). P.F. was supported by the Czech Science Foundation (project 21-27624S), and the European Union Operational Programme Just Transition (LERCO project CZ.10.03.01/00/22_003/0000003). We acknowledge support from Polish scientific project grant NCN OPUS 2015/17/B/HS3/01327, and the Russian Science Foundation grants 21-18-00026 to N.I.S. and 22-18-00470 to A.A.T. We acknowledge support from the Museum of the Institute of Plant and Animal Ecology (UB RAS) to P. Kosintsev. K.N.S. was supported by grant FWRZ-2021-0006. This study depended on support from the research computing group at Harvard Medical School. The ancient-DNA data generation and analysis was supported by the National Institutes of Health (R01-HG012287), the John Templeton Foundation (grant 61220), a gift from J.-F. Clin, the Allen Discovery Center programme, a Paul G. Allen Frontiers Group advised programme of the Paul G. Allen Family Foundation, and the Howard Hughes Medical Institute (to D.R.). The accepted version of this article (before the editing, proofreading and formatting changes following the paper being accepted) is subject to the Howard Hughes Medical Institute (HHMI) Open Access to Publications policy; HHMI lab heads have previously granted a non-exclusive CC BY 4.0 licence to the public and a sublicensable licence to HHMI in their research articles. Pursuant to those licences, the accepted manuscript can be made freely available under a CC BY 4.0 licence immediately on publication.

Author contributions I.L., N.P., D.A., L.V. and D.R. wrote the manuscript. A.S.-N., P.F.P., S.M., N.R., R.P. and D.R. supervised parts of the study. I.L. and N.P. carried out the main genetic analyses. R.F., H.R., I.O. and P.F.P. contributed other genetic analyses. D.A. and L.V. edited archaeological information. D.A., A.A.K., E.P.K., N.I.S., S.C.A., E.B., Z.B., A.B., P.C., A.A.C., I.C., M. Constantinescu, M. Csányi, J.D., S.É., A. Faifert, P.F., A. Frînculeasa, M.N.F., T.Ha., T.Hi, P. Jelínek,

V.I.K., V.K., A. Kitova, A. Korolev, G.K., N.A.N., B.B., P.K.D., P. Jarosz, E.N.K., A.V.K., J.K., P.Ko., P.Ku., R.M., M.Mi., E. Melis, V.M., E. Molnár, J.M., M.N., O.N., M.P.R., M.O.-G., G.P., S.P., T.M.S., N.N.S., A.Š., I.S., V.N.S., A. Simalcsik, K.S., K.N.S., J.T., A.A.T., V.T., S.V., O.F., A.S.-N., D.S.A., A.M.M., S.A.A., R.S.M., and R.P. sampled anthropological remains and/or contributed to the creation of the archaeological supplement. N.I.S. shared previously unpublished radiocarbon dates. A.A., E.S.B., M. Mah, A.M. and S.M. processed bioinformatic data. K.C., F.C., O.C., E.C., L.I., A. Kearns, D.K., A.M.L., M. Michel, J.O., L.Q., J.N.W., F.Z. and N.R. carried out wet laboratory work.

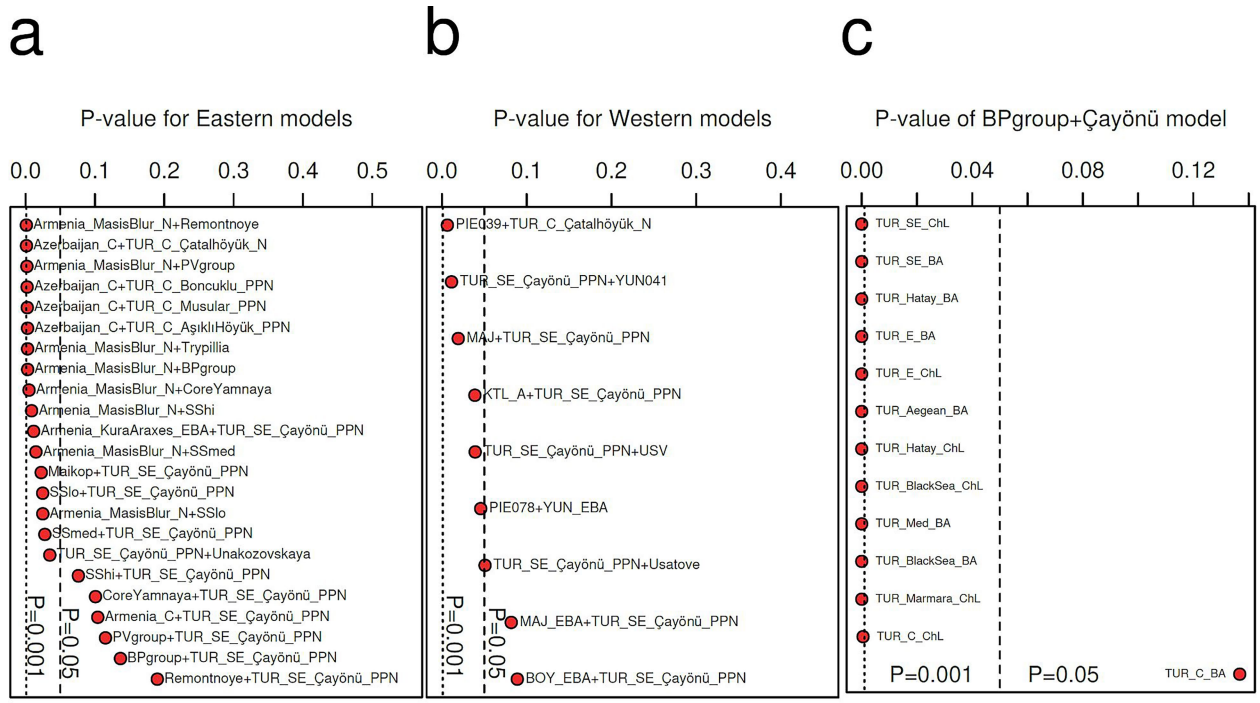
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08531-5.

Correspondence and requests for materials should be addressed to Iosif Lazaridis, Nick Patterson, David Anthony, Leonid Vyazov, Ron Pinhasi or David Reich.

Peer review information *Nature* thanks Kristian Kristiansen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

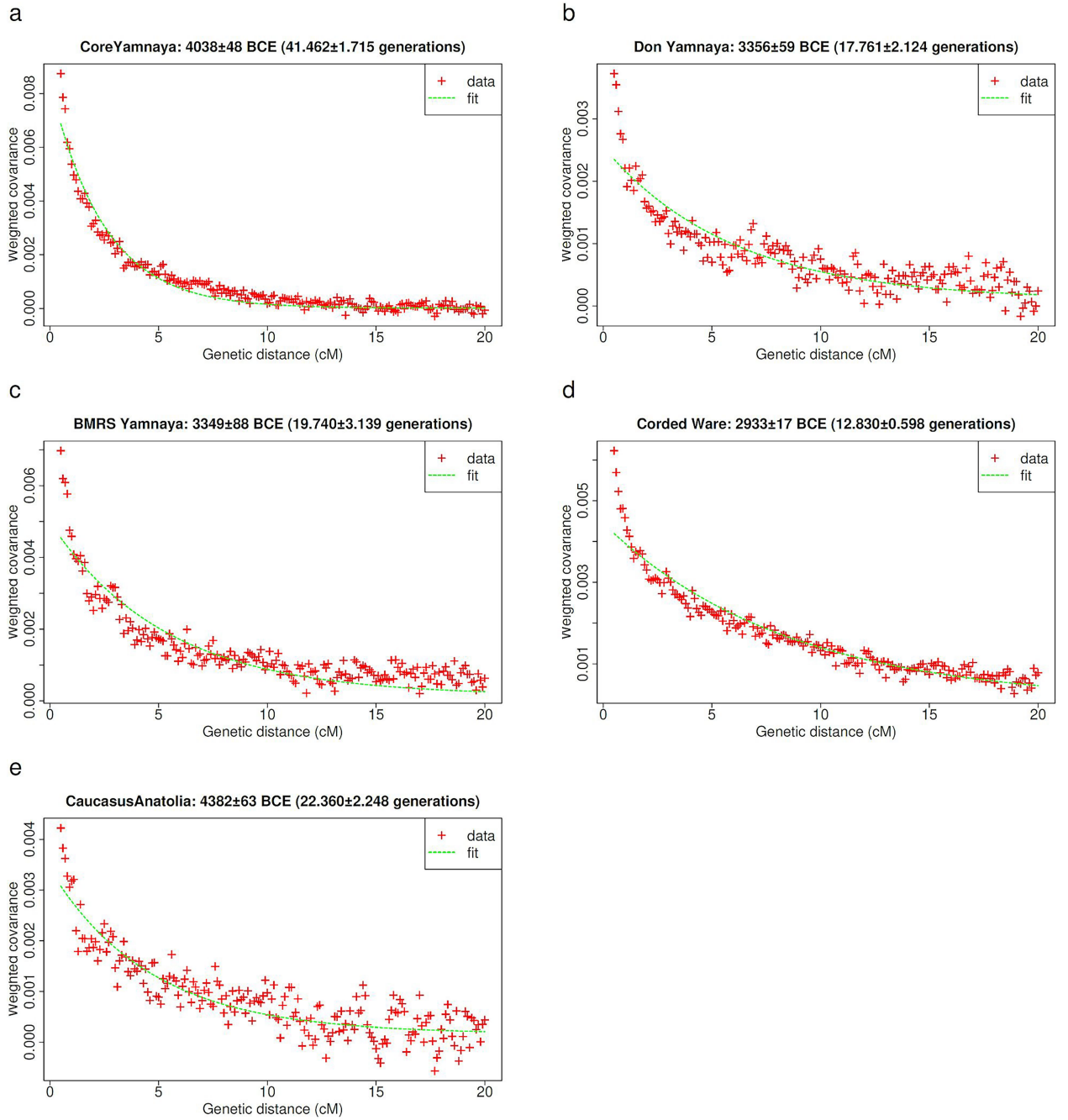


e

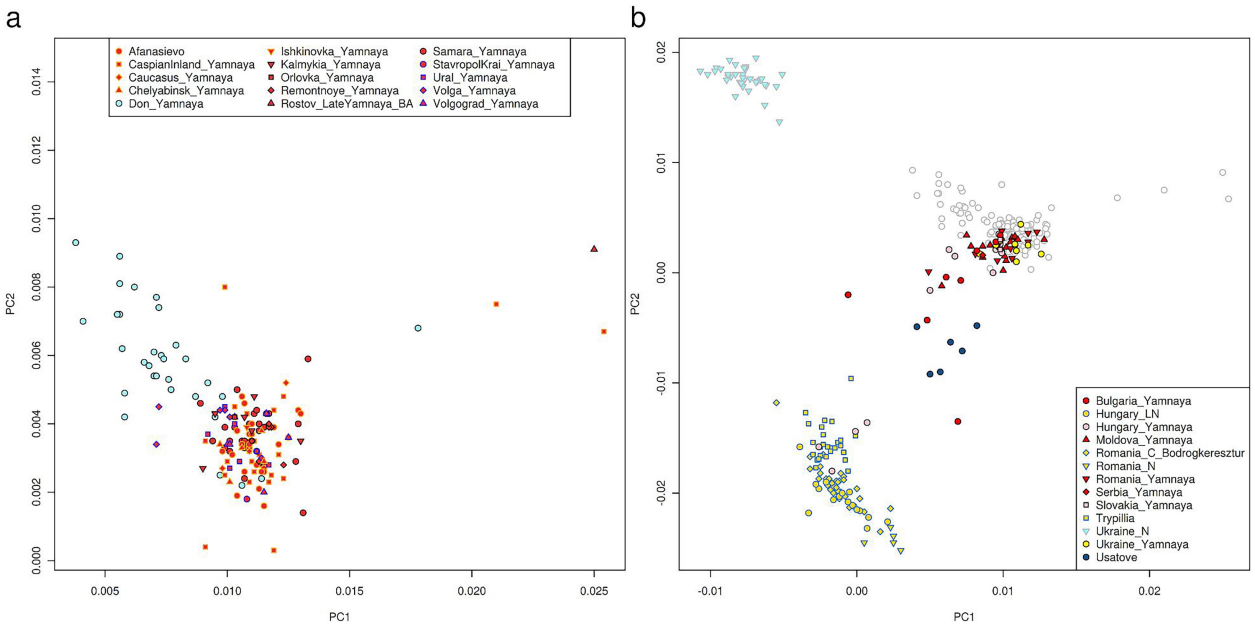
Test	Substratum	P-value	BPgroup	Std. error
Maikop	TUR_SE_Çayönü_PPN	1.07E-36	34.4%	1.8%
Maikop	Armenia_MasisBlur_N	1.02E-08	26.9%	2.3%
Maikop	**Armenia_Aknashen_N**	**1.70E-01**	**14.5%**	**2.9%**
Armenia_C	TUR_SE_Çayönü_PPN	6.32E-04	31.1%	1.8%
Armenia_C	**Armenia_MasisBlur_N**	**7.82E-01**	**27.2%**	**2.2%**
Armenia_C	Armenia_Aknashen_N	9.06E-10	12.0%	3.2%
ART027_d	**TUR_SE_Çayönü_PPN**	**1.71E-01**	**16.7%**	**3.6%**
ART027_d	Armenia_Aknashen_N	5.97E-03	-2.9%	5.2%
ART027_d	**Armenia_MasisBlur_N**	**9.97E-01**	**12.4%**	**3.9%**
TUR_C_BA	**TUR_SE_Çayönü_PPN**	**1.37E-01**	**10.8%**	**1.7%**
TUR_C_BA	Armenia_MasisBlur_N	2.92E-03	5.9%	2.4%
TUR_C_BA	Armenia_Aknashen_N	2.10E-17	-15.8%	3.8%

Extended Data Fig. 1 | See next page for caption.

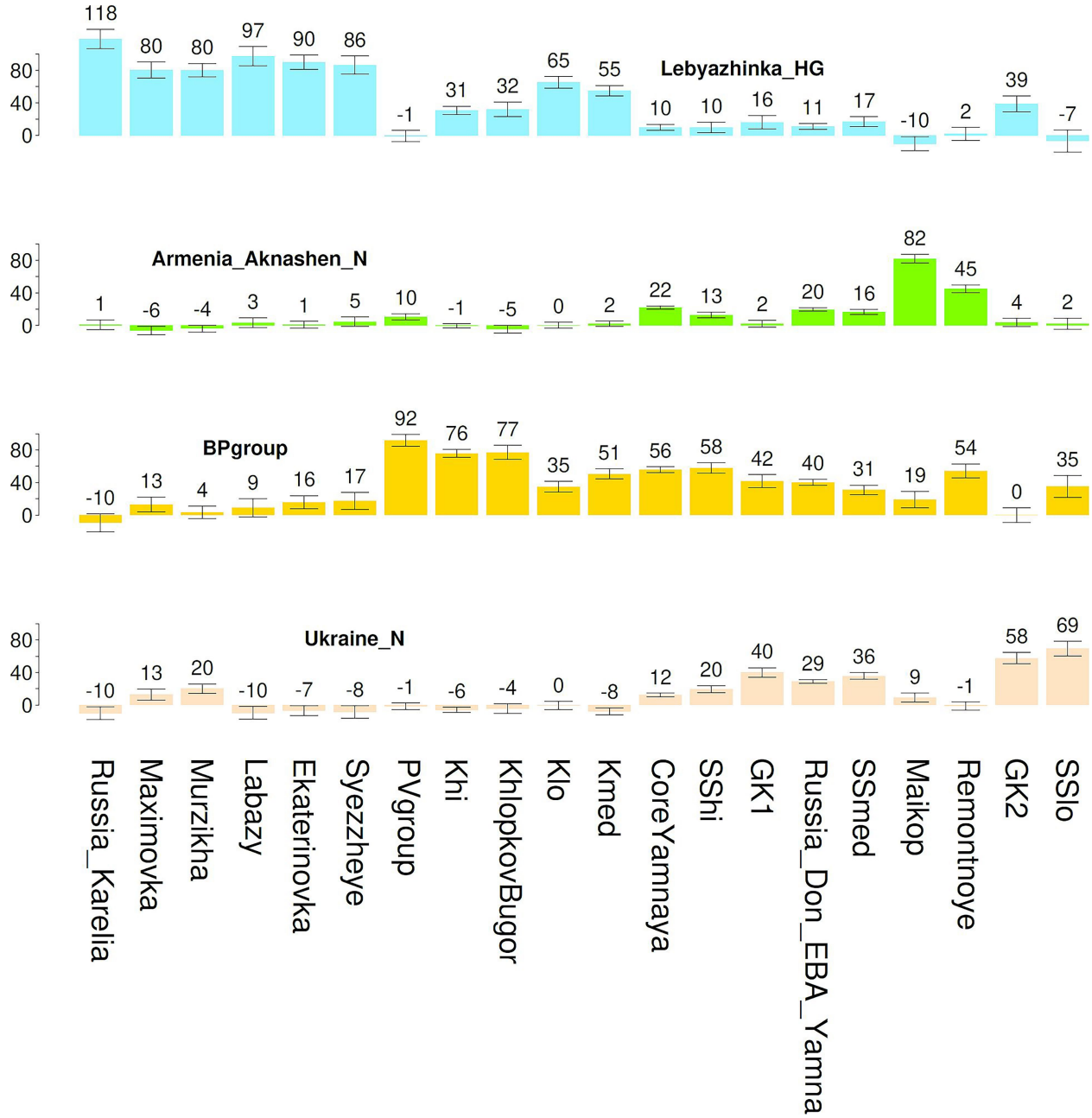
Extended Data Fig. 1 | The origin of Central Anatolian Bronze Age people. **a**, Models with eastern steppe sources (including CLV and Serednii Stih). Fitting models include Mesopotamian (Çayönü) and steppe ancestry. **b**, Models with western sources, including Usatove and those from Southeastern Europe fail except those with Çayönü and either Mayaky or Boyanovo EBA (both of which are Yamnaya-derived). **c**, The steppe (BPgroup)+Çayönü model fails all Chalcolithic/Bronze Anatolians except people of the Central Anatolian Bronze Age. **d**, Steppe (BPgroup) ancestry in the BPgroup+Çayönü model is observed in all individuals of the Central Anatolian Bronze Age (mean and ±3 s.e. estimated by qpAdm are shown for all Chalcolithic and Bronze Age individuals from Anatolia that fit the model at $p > 0.05$) as well as in individual ART027_d from Chalcolithic Arslantepe in Eastern Anatolia. **e**, BPgroup-related ancestry admixed with different substrata: Aknashen-related in the North Caucasus Maikop, Masis Blur-related in Chalcolithic Armenia, and Mesopotamian-related (Çayönü) in the ancestors of the Central Anatolian Bronze Age, following the route (**f**) from the North Caucasus to Anatolia; sites with BPgroup-related ancestry marked in bold. In all panels p-values estimated by qpWave are shown.



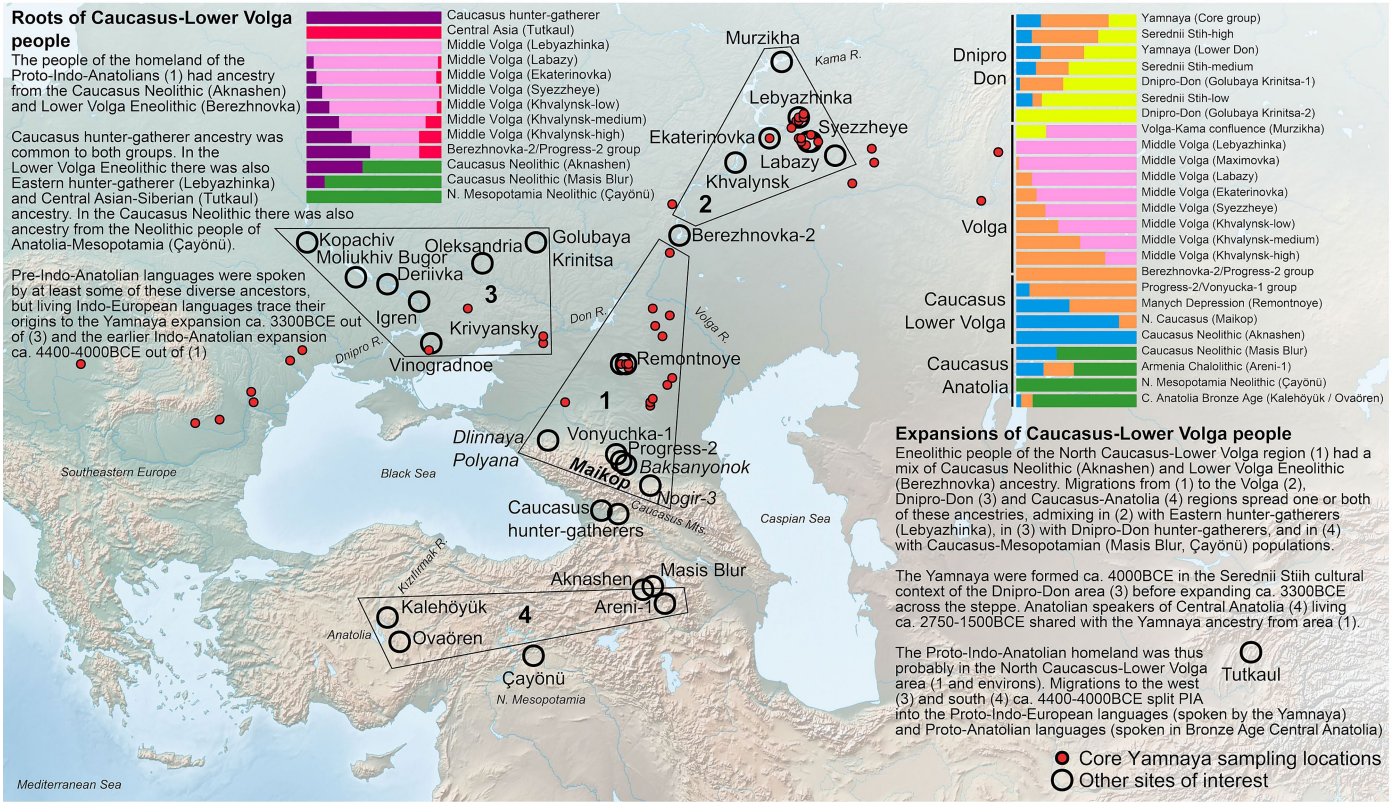
Extended Data Fig. 2 | Admixture date estimates. We estimate admixture dates for the Core Yamnaya as a mixture of European hunter-gatherer and West Asian populations (**a**), for the Don Yamnaya as a mixture of Core Yamnaya and UNHG (**b**), for the Bulgaria, Moldova, Romania, and Serbia (BMRS) Yamnaya as a mixture of Core Yamnaya and European Neolithic/Chalcolithic farmers (**c**), for the Corded Ware as a mixture of Core Yamnaya and Globula Amphora (**d**), and for a combined Caucasus-Anatolia population (Maikop-Armenia_C-TUR_C_BA) a mixture of European hunter-gatherer and West Asian populations which occurred ca. 4400 BC (**e**). The Core Yamnaya were formed ca. 4000 BC, followed by admixture ca. 3350 BC with UNHG and European farmers in the east and west of the Dnipro-Don region and <3000 BC in central-eastern Europe.



Extended Data Fig. 3 | Population structure in people with a Yamnaya cultural affiliation. a, Individuals are projected in the same space as in Fig. 1, showing that the Core Yamnaya cluster (red fill symbols) from diverse sites is differentiated from the Don Yamnaya (blue fill) who tend towards the UNHG. **b**, Yamnaya individuals in the West (Ukraine, Hungary, Slovakia, and Southeastern Europe) include a tight cluster of individuals as well as others that tend towards the direction of European Neolithic and Chalcolithic groups from Romania and Hungary. Individuals from Russia are shown in grey circles in panel **b**. Coordinates of plotted points can be found in Supplementary Table 6.



Extended Data Fig. 4 | A 4-way model for the entire Dnipro-Don-Volga-Caucasus region. Mean and ±1 standard error estimated by qpAdm is shown.



Extended Data Fig. 5 | The origin of Indo-Anatolian and Indo-European languages. Genetic reconstruction of the ancestry of Pontic-Caspian steppe and West Asian populations points to the North Caucasus-Lower Volga area as the homeland of Indo-Anatolian languages and to the Serednii Stih archaeological culture of the Dnipro-Don area as the homeland of Indo-European languages. The Caucasus-Lower Volga people had diverse distal roots, estimated using the *qpAdm* software on the left barplot, as Caucasus hunter-gatherer (purple), Central Asian (red), Eastern hunter-gatherer (pink), and West Asian Neolithic (green). Caucasus-Lower Volga expansions, estimated using *qpAdm* on the right barplot, disseminated Caucasus Neolithic (blue)-Lower Volga Eneolithic (orange) proximal ancestries, mixing with the inhabitants of the North Pontic region (yellow), Volga region (yellow), and West Asia (green).

Extended Data Table 1 | F_{ST} values among select populations of the Dnipro, Don, Volga, and Caucasus areas

	BPgroup	CoreYamnaya	Ekaterinovka	GK1	Khi	KhlopkovBugor	Klo	Kmed	Labazy	Maikop	Maximovka	Murzikha	PVgroup	Remontnoye	Russia_Caucasus_LateMaikop	Russia_Don_EBA_Yamnaya	SShi	SSmed	Syezzheye	Ukraine_N	Unakozovskaya	UpperVolga
BPgroup		0.001	0.001	0.003	0.001	0.002	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.005	0.001	0.001	0.001	0.003	0.001	0.003	0.001
CoreYamnaya	0.011		0.000	0.003	0.000	0.002	0.000	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.004	0.000	0.001	0.001	0.002	0.001	0.002	0.000
Ekaterinovka	0.030	0.032		0.003	0.001	0.002	0.000	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.004	0.000	0.001	0.001	0.003	0.001	0.003	0.000
GK1	0.042	0.041	0.045		0.003	0.007	0.003	0.003	0.005	0.004	0.006	0.003	0.005	0.005	0.018	0.003	0.004	0.005	0.009	0.003	0.006	0.003
Khi	0.007	0.014	0.019	0.039		0.002	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.004	0.001	0.001	0.001	0.003	0.001	0.003	0.001
KhlopkovBugor	0.010	0.017	0.022	0.037	0.008		0.002	0.002	0.003	0.003	0.003	0.002	0.003	0.003	0.009	0.002	0.003	0.003	0.005	0.002	0.004	0.002
Klo	0.018	0.022	0.008	0.041	0.009	0.013		0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.004	0.001	0.001	0.001	0.003	0.001	0.003	0.001
Kmed	0.014	0.018	0.015	0.042	0.006	-0.002	0.002		0.002	0.001	0.002	0.001	0.002	0.002	0.005	0.001	0.001	0.001	0.003	0.001	0.003	0.001
Labazy	0.032	0.034	0.009	0.048	0.021	0.027	0.010	0.016		0.002	0.003	0.002	0.003	0.003	0.007	0.002	0.002	0.003	0.004	0.002	0.004	0.002
Maikop	0.031	0.025	0.064	0.064	0.037	0.043	0.052	0.045	0.067		0.003	0.001	0.002	0.002	0.006	0.001	0.002	0.002	0.003	0.001	0.003	0.001
Maximovka	0.044	0.041	0.021	0.048	0.033	0.033	0.021	0.028	0.021	0.076		0.002	0.003	0.003	0.007	0.002	0.003	0.003	0.004	0.002	0.003	0.002
Murzikha	0.056	0.053	0.034	0.065	0.044	0.047	0.034	0.039	0.034	0.088	0.018		0.002	0.002	0.004	0.001	0.001	0.001	0.003	0.001	0.003	0.001
PVgroup	-0.002	0.012	0.035	0.046	0.010	0.012	0.024	0.018	0.038	0.025	0.048	0.061		0.003	0.006	0.002	0.002	0.003	0.004	0.002	0.003	0.002
Remontnoye	0.012	0.011	0.040	0.041	0.015	0.020	0.028	0.024	0.046	0.012	0.052	0.065	0.011		0.006	0.002	0.002	0.002	0.004	0.002	0.003	0.002
Russia_Caucasus_LateMaikop	0.025	0.020	0.058	0.065	0.033	0.037	0.048	0.041	0.059	-0.001	0.063	0.081	0.026	0.002		0.004	0.006	0.007	0.011	0.004	0.007	0.004
Russia_Don_EBA_Yamnaya	0.014	0.005	0.029	0.040	0.014	0.019	0.019	0.018	0.030	0.030	0.037	0.048	0.016	0.015	0.025		0.001	0.001	0.003	0.001	0.003	0.001
SShi	0.009	0.011	0.027	0.034	0.013	0.014	0.017	0.017	0.029	0.030	0.036	0.048	0.010	0.016	0.034	0.011		0.002	0.004	0.001	0.003	0.001
SSmed	0.011	0.010	0.021	0.034	0.011	0.012	0.015	0.015	0.021	0.030	0.030	0.041	0.013	0.014	0.019	0.008	0.004		0.004	0.001	0.003	0.001
Syezzheye	0.045	0.047	0.022	0.059	0.034	0.035	0.026	0.033	0.029	0.082	0.043	0.050	0.049	0.056	0.077	0.042	0.040	0.034		0.003	0.004	0.003
Ukraine_N	0.046	0.039	0.036	0.047	0.040	0.042	0.032	0.037	0.036	0.063	0.038	0.048	0.049	0.049	0.055	0.029	0.031	0.017	0.055		0.003	0.001
Unakozovskaya	0.059	0.057	0.094	0.090	0.068	0.069	0.083	0.076	0.096	0.034	0.107	0.117	0.058	0.039	0.030	0.060	0.062	0.061	0.107	0.092		0.003
UpperVolga	0.044	0.040	0.021	0.048	0.033	0.035	0.019	0.028	0.019	0.073	0.015	0.027	0.049	0.051	0.067	0.033	0.035	0.026	0.038	0.029	0.103	

F_{ST} values are shown below the diagonal and their standard errors above it.

Extended Data Table 2 | Extraordinary genetic homogeneity in the Core Yamnaya

Population	P-value	Included in Core Yamnaya	Total individuals
Populations that are a clade with Samara Yamnaya			
China_Xinjiang_G218_BA_Afanasievo_oWestEurasia	9.7E-01	1	1
Russia_Chelyabinsk_EBA_Yamnaya	9.5E-01	5	5
Russia_Volgograd_EBA_Yamnaya	9.0E-01	3	5
Russia_Ural_EBA_Yamnaya_contam	8.2E-01	0	1
Usatove_EBA	7.9E-01	0	1
Russia_Ural_EBA_Yamnaya	7.3E-01	5	7
Russia_Afanasievo_Yenisei	6.7E-01	1	1
Russia_MBA_Poltavka	6.5E-01	0	6
Romania_EBA_Catacomb	6.3E-01	0	2
Russia_Orlovka_EBA_Yamnaya	5.5E-01	1	1
Ukraine_MBA	5.1E-01	0	1
Russia_Samara_EBA_Yamnaya_possible	5.0E-01	0	1
Kazakhstan_EBA_Yamnaya.SG	4.8E-01	1	1
Ukraine_EBA_Yamnaya_contam	4.4E-01	0	1
Russia_LowerVolga_EBA_Yamnaya	3.9E-01	0	1
Moldova_Crasnoe_Eneolithic	3.9E-01	0	1
Russia_EBA_o1.SG	3.6E-01	0	1
Ukraine_EBA_Catacomb	3.5E-01	0	2
Ukraine_MBA_Catacomb_o1	3.4E-01	0	1
Moldova_Eneolithic	3.3E-01	0	1
Russia_BA_WestManych_Catacomb	3.2E-01	0	1
Mongolia_Chalcolithic_Afanasievo_1	2.9E-01	0	1
Russia_Kalmykia_EBA	2.8E-01	0	1
Russia_Afanasievo.SG	2.8E-01	0	2
Russia_UpperOb_Eneolithic_Afanasievo	2.8E-01	6	6
Russia_Volgograd_EBA_Yamnaya_o	2.8E-01	0	1
Russia_Ishkinovka_EBA_Yamnaya	2.5E-01	1	1
Usatove_Yamnaya	2.1E-01	0	1
Latvia_LN_CordedWare	1.7E-01	0	1
Hungary_EBA_Yamnaya_1drel.I3510_contam	1.7E-01	0	1
Brillenhohle.pmd	1.6E-01	0	1
Russia_Steppe_Catacomb	1.6E-01	0	4
Russia_Volga_EBA_Yamnaya	1.3E-01	4	5
Russia_Kalmykia_EasternManych_EMBA	1.3E-01	0	2
Russia_N_BA_possible	1.2E-01	0	1
Russia_Afanasievo	1.2E-01	18	29
Moldova_Eneolithic_Suvorove	1.1E-01	0	1
Russia_Afanasievo_Khakassia_possible	9.0E-02	0	1
BOY_EBA	8.8E-02	0	5
Russia_Rostov_Steppe_NorthCaucasus_BA	7.4E-02	0	1
Russia_LowerDon_EBA_Yamnaya	6.6E-02	0	1
Moldova_EBA_Yamnaya	6.5E-02	4	16
Ukraine_EBA_Catacomb.SG	6.2E-02	0	1
Russia_Afanasievo_contam	5.8E-02	0	2
Romania_Brailita_EBA_Yamnaya	5.8E-02	0	1
Slovakia_EBA_Yamnaya	5.1E-02	0	2
Ukraine_EBA_Yamnaya	5.1E-02	4	9
Populations that are not a clade with Samara Yamnaya but include at least one individual that is			
Romania_EBA_Yamnaya	3.9E-02	2	8
Russia_Remontnoye_EBA_Yamnaya	3.5E-02	5	6
Russia_Kalmykia_EBA_Yamnaya.SG	1.8E-02	2	6
Russia_Caucasus_EBA_Yamnaya	1.6E-02	1	3
Hungary_EBA_Yamnaya	1.5E-04	1	5
Russia_CaspianInland_EBA_Yamnaya	1.2E-04	12	26
Russia_UpperYenisey_Eneolithic_Afanasievo	2.3E-05	1	4
Russia_Don_EBA_Yamnaya	2.8E-50	2	23

We tested all populations and individuals for cladality with Samara Yamnaya. We list populations for which this is not rejected (qpWave $p > 0.05$) and populations that include individuals that fit Core Yamnaya selection criteria (qpWave $p > 0.2$, at least 300k SNPs, and Yamnaya or Afanasievo culture).

Extended Data Table 3 | F_{ST} values among populations that include Core Yamnaya individuals

	Hungary_EBA_Yamnaya	Moldova_EBA_Yamnaya	Romania_EBA_Yamnaya	Russia_Afanasievo	Russia_CaspianInland_EBA_Yamnaya	Russia_Caucasus_EBA_Yamnaya	Russia_Chelyabinsk_EBA_Yamnaya	Russia_Don_EBA_Yamnaya	Russia_Kalmykia_EBA_Yamnaya.SG	Russia_Remontnoye_EBA_Yamnaya	Russia_Samara_EBA_Yamnaya	Russia_UpperOb_Eneolithic_Afanasievo	Russia_UpperYenisey_Eneolithic_Afanasievo	Russia_Ural_EBA_Yamnaya	Russia_Volga_EBA_Yamnaya	Russia_Volgograd_EBA_Yamnaya	Ukraine_EBA_Yamnaya
Hungary_EBA_Yamnaya		0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Moldova_EBA_Yamnaya	0.001		0.001	0.000	0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Romania_EBA_Yamnaya	0.001	0.001		0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Afanasievo	0.006	0.004	0.005		0.000	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Russia_CaspianInland_EBA_Yamnaya	0.004	0.003	0.002	0.006		0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Caucasus_EBA_Yamnaya	0.001	0.002	0.001	0.003	0.003		0.002	0.001	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.002
Russia_Chelyabinsk_EBA_Yamnaya	0.008	0.009	0.009	0.010	0.009	0.009		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Don_EBA_Yamnaya	0.006	0.005	0.006	0.008	0.006	0.005	0.012		0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Kalmykia_EBA_Yamnaya.SG	0.007	0.005	0.004	0.005	0.001	0.004	0.011	0.007		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Remontnoye_EBA_Yamnaya	0.004	0.004	0.003	0.004	0.000	0.003	0.010	0.006	-0.049		0.001	0.001	0.001	0.001	0.001	0.001	0.001
Russia_Samara_EBA_Yamnaya	0.003	0.002	0.003	0.005	0.003	0.003	0.008	0.005	0.005	0.004		0.001	0.001	0.001	0.001	0.001	0.001
Russia_UpperOb_Eneolithic_Afanasievo	0.006	0.005	0.004	0.002	0.003	0.006	0.010	0.008	0.001	0.003	0.004		0.001	0.001	0.001	0.001	0.001
Russia_UpperYenisey_Eneolithic_Afanasievo	0.011	0.010	0.008	0.009	0.009	0.009	0.015	0.012	0.009	0.007	0.010	0.006		0.001	0.001	0.001	0.001
Russia_Ural_EBA_Yamnaya	0.002	0.002	0.001	0.004	0.003	0.003	0.006	0.005	0.004	0.003	0.001	0.003	0.008		0.001	0.001	0.001
Russia_Volga_EBA_Yamnaya	0.004	0.004	0.003	0.005	0.005	0.005	0.007	0.007	0.008	0.007	0.003	0.007	0.011	0.003		0.001	0.001
Russia_Volgograd_EBA_Yamnaya	0.005	0.003	0.004	0.007	0.005	0.004	0.009	0.007	0.007	0.005	0.004	0.007	0.009	0.003	0.006		0.001
Ukraine_EBA_Yamnaya	0.003	0.001	0.001	0.004	0.002	0.002	0.008	0.005	0.003	0.003	0.002	0.004	0.009	0.001	0.004	0.004	

F_{ST} values are shown below the diagonal and their standard errors above it.

Extended Data Table 4 | *qpAdm* models that fit non-Core Yamnaya

Modeled group	A	B	P-value	A	B	S.E.
Bulgaria_C_ProtoYamnaya	Serbia_IronGates_Mesolithic	Usatove	0.062	17.5%	82.5%	2.4%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.883	85.4%	14.6%	1.4%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Romania_N	0.682	86.7%	13.3%	1.3%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Trypillia	0.719	82.9%	17.1%	1.7%
Bulgaria_EBA_Yamnaya	CoreYamnaya	Usatove	0.396	66.7%	33.3%	3.3%
Bulgaria_Riltsi_EBA_Yamnaya	Romania_N	Usatove	0.079	25.3%	74.7%	5.8%
Hungary_LateC_EBA_Baden_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.113	3.4%	96.6%	1.5%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Serbia_IronGates_Mesolithic	0.309	95.9%	4.1%	1.4%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Trypillia	0.114	77.6%	22.4%	9.7%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Ukraine_N	0.197	96.2%	3.8%	1.5%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_C_Bodrogkeresztur	Usatove	0.099	94.2%	5.8%	2.5%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_N	Serbia_IronGates_Mesolithic	0.323	87.0%	13.0%	1.7%
Hungary_LateC_EBA_Baden_Yamnaya	Romania_N	Ukraine_N	0.094	86.9%	13.1%	1.7%
Moldova_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.724	93.8%	6.2%	0.9%
Moldova_EBA_Yamnaya	CoreYamnaya	Romania_N	0.571	94.3%	5.7%	0.8%
Moldova_EBA_Yamnaya	CoreYamnaya	Trypillia	0.675	92.7%	7.3%	1.0%
Moldova_EBA_Yamnaya	CoreYamnaya	Usatove	0.367	86.0%	14.0%	2.2%
Moldova_GlobularAmphora_Yamnaya	CoreYamnaya	Trypillia	0.053	88.7%	11.3%	2.6%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.561	91.6%	8.4%	2.4%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Romania_N	0.524	92.4%	7.6%	2.1%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Trypillia	0.601	90.1%	9.9%	2.8%
Romania_Brailita_EBA_Yamnaya	CoreYamnaya	Usatove	0.509	81.0%	19.0%	5.6%
Romania_EBA_Yamnaya	CoreYamnaya	Romania_N	0.096	95.8%	4.2%	1.0%
Romania_EBA_Yamnaya	CoreYamnaya	Usatove	0.143	89.1%	10.9%	2.5%
Serbia_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.196	87.3%	12.7%	2.0%
Serbia_EBA_Yamnaya	CoreYamnaya	Romania_N	0.097	88.6%	11.4%	1.9%
Serbia_EBA_Yamnaya	CoreYamnaya	Trypillia	0.200	85.1%	14.9%	2.3%
Ukraine_EBA_Yamnaya	CoreYamnaya	Romania_C_Bodrogkeresztur	0.561	93.3%	6.7%	1.3%
Ukraine_EBA_Yamnaya	CoreYamnaya	Romania_N	0.481	94.0%	6.0%	1.2%
Ukraine_EBA_Yamnaya	CoreYamnaya	Trypillia	0.756	92.0%	8.0%	1.5%
Ukraine_EBA_Yamnaya	CoreYamnaya	Usatove	0.534	84.5%	15.5%	3.1%

We use the following sources to model Yamnaya-related populations other than the Core and Don Yamnaya: CoreYamnaya, Romania_C_Bodrogkeresztur, Romania_N, Serbia_IronGates_Mesolithic, Trypillia, Ukraine_N, Usatove. The Baden individuals from Hungary represent a reburial into a kurgan[54] and are predominantly of European farmer, not Yamnaya, ancestry. The Riltsi individual is shown with Usatove ancestry here and can also be modeled with about half Remontnoye ancestry, as the Usatove have ancestry from the CLV cline[17].

Extended Data Table 5 | Cross-regional shared identity-by-descent (IBD) segments

Individual 1	Individual 2	Group 1	Group 2	Segment length (cM)
I22201	I1924	BPgroup	SShi	35.8
I22202	I6734	BPgroup	Khi	32.1
I1634	I22199	Armenia_C	BPgroup	31.4
I6300_enhanced	I22202	KhlopkovBugor	BPgroup	22.0
I6406	I22200	Kmed	BPgroup	20.1
PG2004	I11837	BPgroup	Khi	18.4
I6301_enhanced	I22199	KhlopkovBugor	BPgroup	18.2
I6301_enhanced	PG2001	KhlopkovBugor	PVgroup	17.6
I28683	PG2004	Remontnoye	BPgroup	16.6
I10567	I28682	Russia_CaspianInland_EBA_Yamnaya	Remontnoye	16.2
PG2001	I3950	PVgroup	Russia_Afanasievo	15.9
PG2001	I6062	PVgroup	Ekaterinovka	15.9
I22199	I8282	BPgroup	Ekaterinovka	15.8
I22201	I10208	BPgroup	Moldova_EBA_Yamnaya	15.5
I1924	I20188	SShi	Klo	15.4
I32501	I8448	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	15.4
I12637	I8457	Moldova_EBA_Yamnaya	Murzikha	15.4
I32821	I8449	Russia_UpperOb_Eneolithic_Afanasievo	Murzikha	15.4
MA2213_wNonUDG.SG	VJ1001	TUR_C_Ovaören_EBA	PVgroup	15.2
I32501	I8455	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	15.2
I6301_enhanced	I22199	KhlopkovBugor	BPgroup	14.9
I8411_enhanced	I26785	UpperVolga	Russia_Don_EBA_Yamnaya	14.9
I22201	I1924	BPgroup	SShi	14.8
I22199	I28682	BPgroup	Remontnoye	14.8
I0122	I22202	Klo	BPgroup	14.6
I32501	I8454	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	14.5
I22199	I6734	BPgroup	Khi	14.5
I22201	I11752	BPgroup	Russia_Afanasievo	14.3
I6064	I22199	Ekaterinovka	BPgroup	14.2
I0122	I22199	Klo	BPgroup	14.2
I1634	I1924	Armenia_C	SShi	13.9
I6301_enhanced	I22201	KhlopkovBugor	BPgroup	13.9
I6918	I8446	Russia_Volgograd_EBA_Yamnaya	Maximovka	13.9
I22202	I6739	BPgroup	Khi	13.9
PG2004	I23651	BPgroup	Ekaterinovka	13.7
I0357	I11842	Russia_Samara_EBA_Yamnaya	Murzikha	13.7
I22202	I3952	BPgroup	Russia_Afanasievo	13.7
I0122	I20190	Klo	Russia_Samara_EBA_Yamnaya	13.6
I8951	I11842	Russia_Don_EBA_Yamnaya	Murzikha	13.5
PG2004	I8290	BPgroup	Ekaterinovka	13.4
I0231	I8456	Russia_Samara_EBA_Yamnaya	Murzikha	13.4
I25159	I22199	Russia_Afanasievo	BPgroup	13.3
I4111	I6109	Ukraine_N	Klo	13.3
I22199	I26787	BPgroup	Russia_Don_EBA_Yamnaya	13.3
I6301_enhanced	PG2004	KhlopkovBugor	BPgroup	12.9
I8449	I2105	Murzikha	Ukraine_EBA_Yamnaya	12.9
I20189	I22200	Ekaterinovka	BPgroup	12.8
I6297	I22201	Russia_Orlovka_EBA_Yamnaya	BPgroup	12.8
I6705	I28682	Russia_Samara_EBA_Yamnaya	Remontnoye	12.8
I32821	I22200	Russia_UpperOb_Eneolithic_Afanasievo	BPgroup	12.7
I32501	I8449	Russia_UpperYenisey_Eneolithic_Afanasievo	Murzikha	12.6
I22201	I6739	BPgroup	Khi	12.4
I0231	I28682	Russia_Samara_EBA_Yamnaya	Remontnoye	12.3
PG2004	I6739	BPgroup	Khi	12.3
I6918	I22200	Russia_Volgograd_EBA_Yamnaya	BPgroup	12.3
I22201	I3952	BPgroup	Russia_Afanasievo	12.2
I6406	I1450	Kmed	Russia_Samara_EBA_Yamnaya	12.2
I22199	I5273	BPgroup	Russia_Afanasievo	12.1
I4114	I12964	Ukraine_N	UpperVolga	12.1
I11838	I23651	Russia_Volga_EBA_Yamnaya	Ekaterinovka	12.0
I6907	I11841	Russia_Samara_EBA_Yamnaya	Murzikha	12.0
I22201	I1924	BPgroup	SShi	12.0

We list all segments≥12 cM shared between individuals from two different regions defined as follows. "Dnipro cline": CoreYamnaya, GK1, GK2, Russia_Don_EBA_Yamnaya, SShi, SSlo, SSmed, Ukraine_N. Volga River basin ancestry gradients (downriver "Volga Cline" and upriver "European Hunter-Gatherer Cline"): Ekaterinovka, Khi, KhlopkovBugor, Klo, Kmed, Labazy, Lebyazhinka_HG, Maximovka, Murzikha, Syezzheye, UpperVolga. "Caucasus-Lower Volga Eneolithic": BPgroup, PVgroup. "CLV-South": Remontnoye, Maikop, Unakozovskaya, Armenia_C, TUR_C_Kalehöyük_MLBA, TUR_C_Ovaören_EBA.