

Current Biology

The genetic origin of Huns, Avars, and conquering Hungarians

Highlights

- 265 new ancient genomes help to unravel the origin of migration-period populations
- Genetic continuity is detected between Xiongnu and European Huns
- European Avars most likely originated from Mongolia and were related to Huns
- Conquering Hungarians had Ugric ancestry and later admixed with Sarmatians and Huns

Authors

Zoltán Maróti, Endre Neparáczki, Oszkár Schütz, ..., Szilárd Sándor Gál, Péter Tomka, Tibor Török

Correspondence

torokt@bio.u-szeged.hu

In brief

Maróti et al. show that immigrants were in minority compared with the locals of the Carpathian Basin in each period. Several Hun period immigrants had Asian Hun (Xiongnu) ancestry. The Avar immigrant elite had ancient Mongolian origin. Conquering Hungarians and Mansis had common ancestors, but proto-Hungarians further admixed with Sarmatians and Huns.



Article

The genetic origin of Huns, Avars, and conquering Hungarians

Zoltán Maróti,^{1,2} Endre Neparáczki,^{1,3} Oszkár Schütz,³ Kitti Maár,³ Gergely I.B. Varga,¹ Bence Kovács,^{1,3} Tibor Kalmár,² Emil Nyerki,^{1,2} István Nagy,^{4,5} Dóra Latinovics,⁴ Balázs Tihanyi,^{1,6} Antónia Marcsik,⁶ György Pálfi,⁶ Zsolt Bernert,⁷ Zsolt Gallina,^{8,9} Ciprián Horváth,⁹ Sándor Varga,¹⁰ László Költő,¹¹ István Raskó,¹² Péter L. Nagy,¹³ Csilla Balogh,¹⁴ Albert Zink,¹⁵ Frank Maixner,¹⁵ Anders Götherström,¹⁶ Robert George,¹⁶ Csaba Szalontai,¹⁷ Gergely Szenthe,¹⁷ Erwin Gáll,¹⁸ Attila P. Kiss,¹⁹ Bence Gulyás,²⁰ Bernadett Ny. Kovacsóczy,²¹ Szilárd Sándor Gál,²² Péter Tomka,²³ and Tibor Török^{1,3,24,*}

¹Department of Archaeogenetics, Institute of Hungarian Research, 1041 Budapest, Hungary

²Department of Pediatrics and Pediatric Health Center, University of Szeged, 6725 Szeged, Hungary

³Department of Genetics, University of Szeged, 6726 Szeged, Hungary

⁴SeqOmics Biotechnology Ltd., 6782 Mórahalom, Hungary

⁵Institute of Biochemistry, Biological Research Centre, 6726 Szeged, Hungary

⁶Department of Biological Anthropology, University of Szeged, 6726 Szeged, Hungary

⁷Department of Anthropology, Hungarian Natural History Museum, 1083 Budapest, Hungary

⁸Ásatárs Ltd., 6000 Kecskemét, Hungary

⁹Department of Archaeology, Institute of Hungarian Research, 1041 Budapest, Hungary

¹⁰Móra Ferenc Museum, 6720 Szeged, Hungary

¹¹Rippl-Rónai Municipal Museum with Country Scope, 7400 Kaposvár, Hungary

¹²Institute of Genetics, Biological Research Centre, 6726 Szeged, Hungary

¹³Praxis Genomics LLC, Atlanta, GA 30328, USA

¹⁴Department of Art History, Istanbul Medeniyet University, 34720 Istanbul, Turkey

¹⁵Institute for Mummy Studies, EURAC Research, 39100 Bolzano, Italy

¹⁶Department of Archaeology and Classical Studies, Stockholm University, 11418 Stockholm, Sweden

¹⁷Hungarian National Museum, Department of Archaeology, 1088 Budapest, Hungary

¹⁸“Vasile Pârvan” Institute of Archaeology, 010667 Bucharest, Romania

¹⁹Faculty of Humanities and Social Sciences, Institute of Archaeology, Pázmány Péter Catholic University, 1088 Budapest, Hungary

²⁰Institute of Archaeological Sciences, Eötvös Loránd University, 1088 Budapest, Hungary

²¹Katona József Museum, 6000 Kecskemét, Hungary

²²Mureş County Museum, 540088 Târgu Mureş, Romania

²³Department of Archaeology, Rómer Flóris Museum of Art and History, 9021 Győr, Hungary

²⁴Lead contact

*Correspondence: torokt@bio.u-szeged.hu

<https://doi.org/10.1016/j.cub.2022.04.093>

SUMMARY

Huns, Avars, and conquering Hungarians were migration-period nomadic tribal confederations that arrived in three successive waves in the Carpathian Basin between the 5th and 9th centuries. Based on the historical data, each of these groups are thought to have arrived from Asia, although their exact origin and relation to other ancient and modern populations have been debated. Recently, hundreds of ancient genomes were analyzed from Central Asia, Mongolia, and China, from which we aimed to identify putative source populations for the above-mentioned groups. In this study, we have sequenced 9 Hun, 143 Avar, and 113 Hungarian conquest period samples and identified three core populations, representing immigrants from each period with no recent European ancestry. Our results reveal that this “immigrant core” of both Huns and Avars likely originated in present day Mongolia, and their origin can be traced back to Xiongnu (Asian Huns), as suggested by several historians. On the other hand, the “immigrant core” of the conquering Hungarians derived from an earlier admixture of Mansis, early Sarmatians, and descendants of late Xiongnu. We have also shown that a common “proto-Ugric” gene pool appeared in the Bronze Age from the admixture of Mezhevskaya and Nganasan people, supporting genetic and linguistic data. In addition, we detected shared Hun-related ancestry in numerous Avar and Hungarian conquest period genetic outliers, indicating a genetic link between these successive nomadic groups. Aside from the immigrant core groups, we identified that the majority of the individuals from each period were local residents harboring “native European” ancestry.



INTRODUCTION

Successive waves of population migrations associated with the Huns, Avars, and Hungarians or Magyars from Asia to Europe had an enduring impact on the population of the Carpathian Basin.¹ This is most conspicuous in the unique language and ethno-cultural traditions of the Hungarians, the closest parallels of which are found in populations east of the Urals. According to present scientific consensus, these eastern links are solely attributed to the last migrating wave of conquering Hungarians (henceforth shortened as Conquerors), who arrived in the Carpathian Basin at the end of the 9th century CE. On the other hand, medieval Hungarian chronicles, foreign written sources, and Hungarian folk traditions maintain that the origin of Hungarians can be traced back to the European Huns, with subsequent waves of Avars and Conquerors considered kinfolk of the Huns.^{2,3}

Both Huns and Avars founded a multiethnic empire in Eastern Europe centered on the Carpathian Basin. The appearance of Huns in European written sources ca 370 CE was preceded by the disappearance of Xiongnu from Chinese sources.⁴ Likewise, the appearance of Avars in Europe in the sixth century broadly correlates with the collapse of the Rouran Empire.⁵ However, the possible relations between Xiongnu and Huns as well as Rourans and Avars remain largely controversial due to the scarcity of sources.⁶

From the 19th century onward, linguists reached a consensus that the Hungarian language is a member of the Uralic language family, belonging to the Ugric branch with its closest relatives the Mansi and Khanty languages.^{7,8} On this linguistic basis, the Hungarian prehistory was rewritten, and the Conquerors were regarded as descendants of a hypothetical Proto-Ugric people. At the same time, the formerly accepted Hun-Hungarian relations were called into question by source criticism of the medieval chronicles.^{9,10}

Due to the scarceness of bridging literary evidence and the complex archaeological record, an archaeogenetic approach is best suited to provide insights into the origin and biological relationship of ancient populations. To this end, we performed whole genome analysis of European Hun, Avar, and Conquest period individuals from the Carpathian Basin to shed light on the long-debated origin of these groups. The majority of our 271 ancient samples (Data S1A) were collected from the Great Hungarian Plain (Alföld), the westernmost extension of the Eurasian steppe, which provided a favorable environment for the arriving waves of nomadic groups. The overview of archaeological sites and time periods of the studied samples is shown in Figure 1, and a detailed archaeological description of the periods, cemeteries, and individual samples is given in Methods S1. From the studied samples, we report 73 direct accelerator mass spectrometry (AMS) radiocarbon dates, of which 50 are first reported in this paper (Data S2).

In this work, we identified immigrant and local groups from each period and revealed the most likely genetic origin of all individuals.

RESULTS

Genome-wide data were generated for 271 ancient individuals using shotgun sequencing. We obtained genome coverages

ranging from 0.15- to 7.09-fold, with an average of 1.97-fold, and quality control with MapDamage 2.0,¹¹ Schmutzi,¹² and ANGSD¹³ estimated negligible contamination in nearly all samples (Data S1A and S1B). All data were pseudohaploidized by randomly calling SNPs at all positions of the human origins (HOs, 600K SNPs) and the 1240K dataset. Most of the analysis was done with the HO dataset, as relevant modern genomes were available only in this format. We identified 43 kinship relations in our samples with the PCAngsd software¹⁴ (Table S1), and just one of the relatives was included in the subsequent analysis. The new data were merged and coanalyzed with 2,364 ancient (Data S3) and 1,397 modern Eurasian genomes (Table S2). As the studied samples represent three archaeologically distinguishable periods from three consecutive historically documented major migration waves into the Carpathian Basin, we evaluated Hun, Avar, and Conquest period samples separately. In order to group the most similar genomes for population genetic analysis, we clustered our samples together with all published ancient Eurasian genomes, according to their pairwise genetic distances obtained from the first 50 principal component analysis (PCA) dimensions (PC50 clustering; STAR Methods; Data S3).

Most individuals in the study had local European ancestry

We performed PCA by projecting our ancient genomes onto the axes computed from modern Eurasian individuals (Figures 2A and S1). In Figure 2A, many samples from nearly each period project onto modern European populations; moreover, these samples form a South-North cline along the PC2 axis, which we termed the Eur-cline. PC50 clustering identified five genetic clusters within the Eur-cline (Figures 2 and S1B; Data S3), well sequestered along the PC2 axis. We selected representative samples from each cluster based on individual distal qpAdm and grouped them as Eur_Core1 to Eur_Core5, respectively. Eur_Core groups include samples from multiple periods, and they are not considered distinct populations, rather they represent distinct local genome types suitable for subsequent modeling. We also showed that each Eur-cline member can be modeled from the Eur_Core groups (Data S4; summarized in Data S1C).

Eur_Core1 clusters with Langobards from Hungary;¹⁵ Iron Age, Imperial, and Medieval individuals from Italy;¹⁶ and Minoans and Mycenaeans from Greece¹⁷ (Data S3). Eur_Core2, 3, and 4 cluster among others with Langobards¹⁵ and Bronze Age samples from Hungary,^{18,19} the Czech Republic, and Germany,¹⁹ whereas Eur_Core5 clusters with Hungarian Scythians.²⁰

Unsupervised ADMIXTURE analysis revealed a gradient-like shift of genomic components along the Eur-cline (Figure 2B) with increasing “Ancient North Eurasian” (ANE) and “Western Hunter-Gatherer” (WHG) and decreasing “early Iranian farmer” (Iran_N) and “early European farmer” (EU_N) components from South to North. It is also apparent that Eur-cline samples contain negligible Asian (“Nganasan” and “Han”) components. ADMIXTURE also confirms that similar genomes had been present in Europe and the Carpathian Basin before the Migration Period, as Eur_Core1 and 5 have comparable patterns with

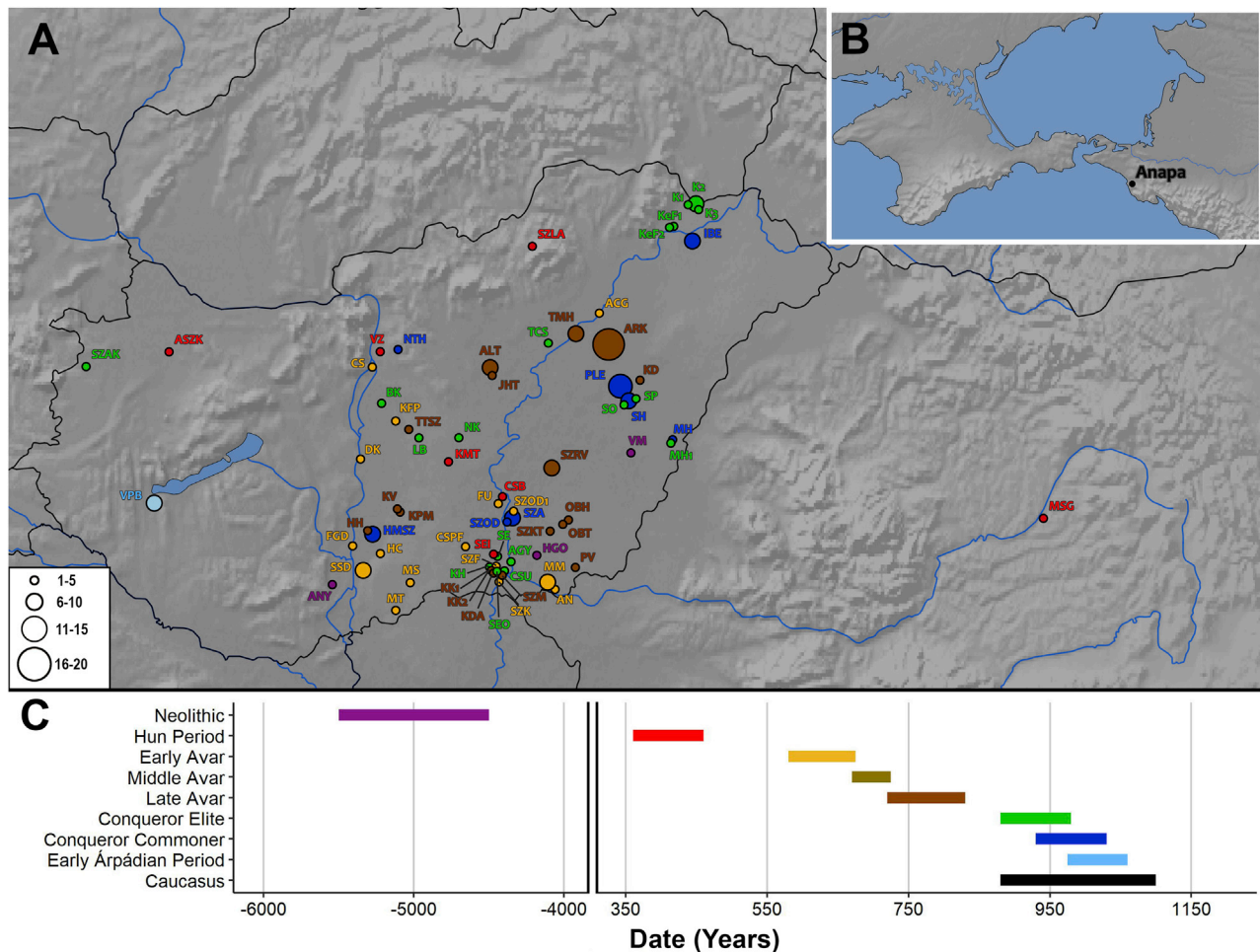


Figure 1. Archaeological sites and time periods of the studied samples

Map of Hungary with neighboring regions and the north Black Sea region (inset) is shown.

(A) Distribution of sites with their associated culture and time period are indicated by color. Color coding corresponds to the time periods labeled with the same color in (C); circle size is proportional to sample size, as indicated.

(B) Inset map of part of the north Black Sea region, from which 3 samples were studied.

(C) Timeline in years of the studied historical periods or archaeological cultures. Negative values indicate BCE dates.

See also [Methods S1](#).

Imperial Period individuals from Italy and Iron Age Scythians from Hungary, respectively.

The diversity of the medieval Hungarian population in the Eur-cline is conspicuous. We considered these groups local residents, although similar populations could hypothetically have been present on the Medieval Pontic Steppe too.

The European Huns had Xiongnu ancestry

Despite the paucity of Hun period samples, we can discern a “Hun-cline” along the PC1 axis (Figure 2A) extending east to west. Two individuals, MSG-1 and VZ-12673 (the same sample as HUN001,²¹ resequenced with higher coverage), project to the extreme eastern pole of the cline, close to modern Kalmyks and Mongols. Both graves belonged to warriors and contained partial horse remains. On PC50 clustering, they tightly cluster together with two other Hun period samples: Kurayly_Hun_380CE (KRY001)²¹ and a Tian Shan Hun outlier

(DA127)²⁰ (Data S3). As latter samples also form a genetic clade with VZ-12673 in qpAdm (see below), we grouped these four samples under the name of Hun_Asia_Core (Figure 2), although analyzed the new samples separately. The Hun_Asia_Core also clusters with numerous Xiongnu, Medieval Mongol, Turk,²² and Xianbei²¹ genomes from Mongolia as well as several Avar samples from this study. ADMIXTURE confirmed the similarity of Hun_Asia_Core individuals and showed prevailing east Eurasian Nganasan and Han components with no traces of WHG (Figure 2B; Data S5), implying that these individuals represent immigrants with no European background.

Outgroup f3-statistics indicated common ancestry of MSG-1 and VZ-12673 (Data S6A), as both individuals shared highest drift with Mongolia-related populations like earlyXiongnu_rest, Ulaanzukh, and SlabGrave.²² Although the error bars overlap with populations related to surrounding regions too, like Russia_MN_Boisman²³ and Tasmola_Korgantas,²¹ f3 results

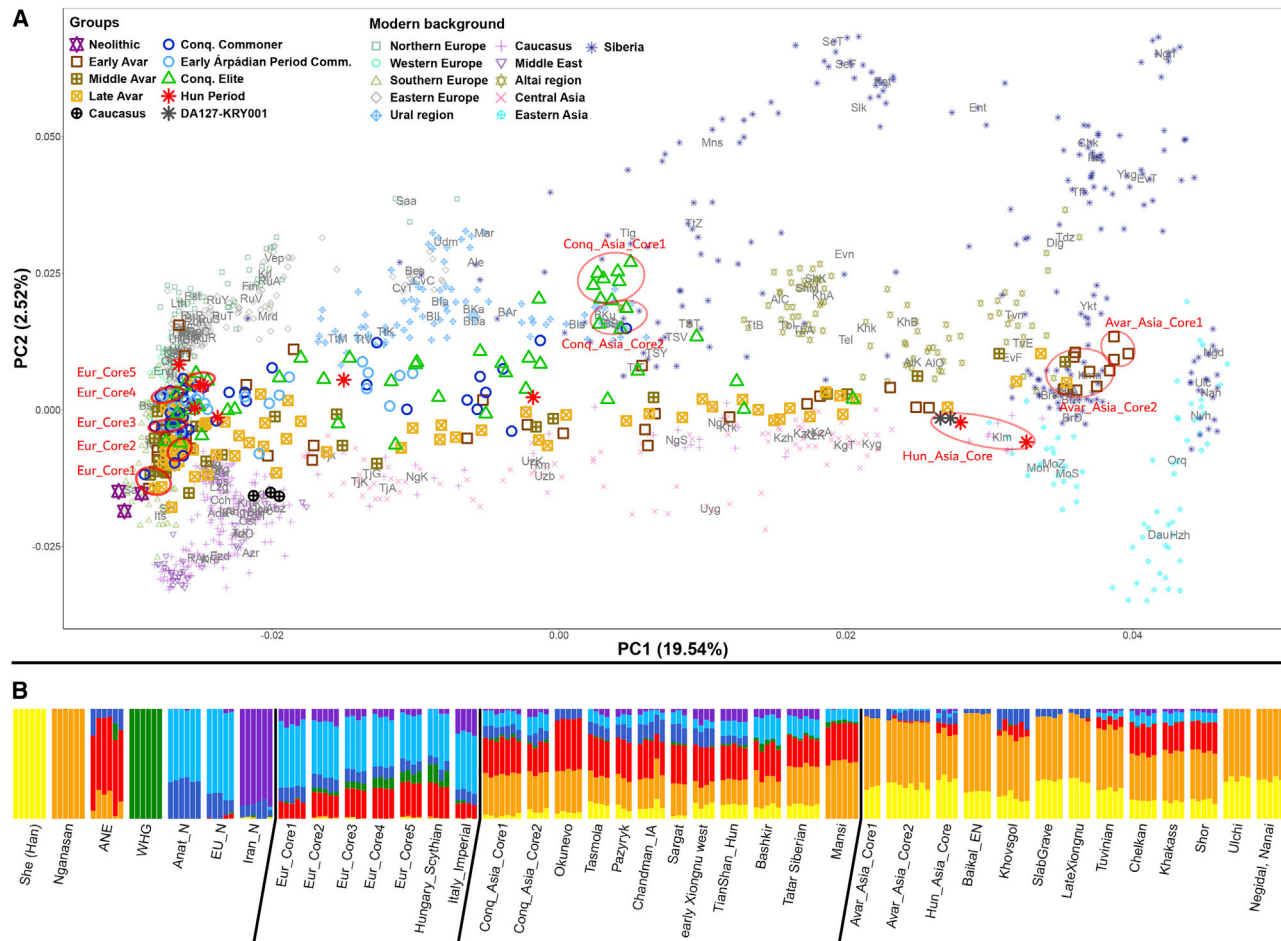


Figure 2. PCA and ADMIXTURE analysis

(A) PCA of 271 ancient individuals projected onto contemporary Eurasians (gray letter codes, defined in Table S2, printed on the median of each population). Labels of modern populations correspond to geographical regions as indicated. Conquest and Avar period samples form two separable genetic clines. Genetically homogeneous groups are encircled with red.

(B) Unsupervised ADMIXTURE (K = 7) results of the red circled core groups and the populations with most similar ADMIXTURE composition to them. The 7 populations representing each ADMIXTURE component are shown at the left.

See also Figure S1, Table S2, and Data S4, S5, S7, S8, and S9.

point to a likely Mongolian origin and early Xiongnu affinity of these individuals.

Distal qpAdm modeling from pre-Iron Age sources indicated major Khovsgol_outlier (DSKC)²² and minor West Liao River Neolithic/Yellow River Late Bronze Age²⁴ ancestries in MSG-1 and VZ-12673 (Figure 3A; Data S7A) predicting Han Chinese admixture in these individuals, whereas proximal modeling from post-Bronze Age sources gave two types of alternative models representing two different time periods (Data S7B). The best p value models showed major late Xiongnu (with Han admixture) and minor Scytho-Siberian/Xianbei ancestries, whereas alternative models indicated major Kazakhstan_OutTianShanHun or Kurayly_Hun_380CE and minor Xiongnu/Xianbei/Han ancestries (Figure 3A). In latter models, VZ-12673 formed a clade with both published Hun_Asia_Core samples. In conclusion, our Hun_Asia_Core individuals could be equally modeled from earlier Xiongnu and later Hun age genomes.

The two other Hun period samples, KMT-2785 and ASZK-1, were located in the middle of the Hun-cline (Figures 2A and S1A), and accordingly, they could be modeled from European and Asian ancestors. The best passing models for KMT-2785 predicted major Late Xiongnu and minor local Eur_Core, whereas alternative model showed major Sarmatian²⁰ and minor Xiongnu ancestries (Data S7C). Both models implicate Sarmatians as in the Late Xiongnus of the first model, and 46%–52% Sarmatian and 48%–54% Ulaanzuukh_SlabGrave components had been predicted.²² The ASZK-1 genome formed a clade with Sarmatians in nearly all models. The rest of the Hun period samples map to the northern half of the Eur-cline; nevertheless, two of these (SEI-1 and SEI-5) could be modeled from major Eur_Core and minor Sarmatian components (Data S7D). The prevalent Sarmatian ancestry in 4 Hun period samples implies significant Sarmatian influence on European Huns (Figure 3B).

CSB-3 was modeled from major Eur_Core and minor Scytho-Siberian ancestries, whereas SEI-6 formed a clade with the

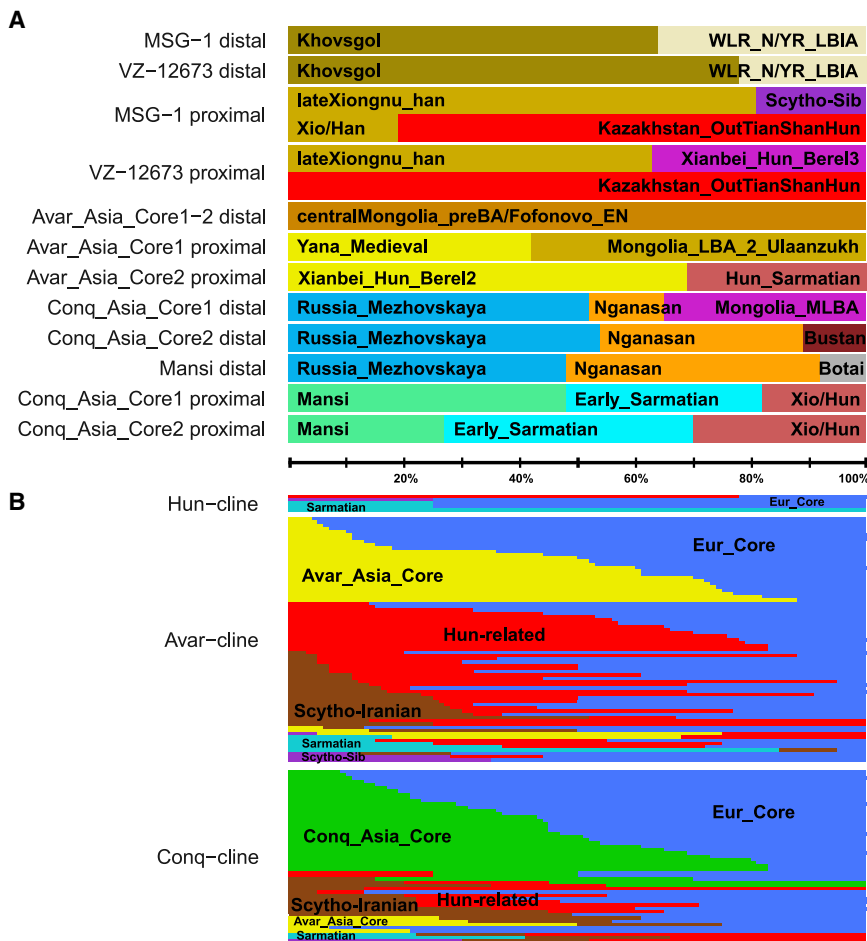


Figure 3. Summary of the qpAdm models

(A) Distal and proximal qpAdm models for Hun_Asia_Core individuals and Avar_Asia_Core and Conq_Asia_Core groups. Distal model of modern Mansi is also shown for comparison.

(B) Proximal qpAdm models for 5 Hun-cline, 53 Conqueror-cline, and 75 Avar-cline individuals. Each individual is represented by a thin column, which are grouped according to the similarity of components. Data are summarized from Data S1C, S7, S8, and S9. Identical or similar genome compositions are shown with similar colors.

and ANE, whereas Iranian and WHG constituents are entirely missing. It follows that Avar_Asia_Core was derived from East Asia, most likely from present day Mongolia.

We performed two-dimensional f4-statistics to detect minor genetic differences within the Avar_Asia_Core group. Avar_Asia_Core individuals could be separated based on their affinity to Bactria-Margiana Archaeological Complex (BMAC) and Steppe Middle-Late Bronze Age (Steppe_MLBA) populations (Figure 4), with 3 individuals bearing negligible proportion of these ancestries. The Steppe_MLBA-ANE f4-statistics gave similar results. As the 3 individuals with the smallest Iranian and Steppe affinities also visibly separated on PCA, we set these apart under the name

Ukraine_Chernyakhiv²⁵ (Eastern Germanic/Goth) genomes (Data S7D). The SZLA-646 outlier individual at the top of the Eur-cline formed a clade with Lithuania_Late_Antiquity²⁰ and England_Saxon²⁶ individuals (Data S7E). The last two individuals presumably belonged to Germanic groups allied with the Huns.

Huns and Avars had related ancestry

Our Avar period samples also form a characteristic PCA “Avar-cline” on Figure 2A, extending from Europe to Asia. PC50 clustering, at level 50, identified a single genetic cluster at the Asian extreme of the cline with 12 samples, derived from 8 different cemeteries, that we termed Avar_Asia_Core (Figure 2; Data S3). In total, 10 of 12 samples of Avar_Asia_Core were assigned to the early Avar period, 4 of them belonging to the elite and 9 of 12 males. Elite status is indicated by richly furnished burials, e.g., swords and sabers with precious metal fittings, gold earrings, gilded belt fittings, etc., as previously described.²⁷

Avar_Asia_Core clusters together with Shamanka_Eneolithic and Lokomotiv_Eneolithic²⁸ samples from the Baikal region, as well as with Mongolia_N_East, Mongolia_N_North,²³ Fofonovo_EN, Ulaanzukh_SlabGrave, and Xiongnu²² from Mongolia (Data S3). This result is recapitulated in ADMIXTURE (Figure 2B), which also shows that Nganasan and Han components are predominant in Avar_Asia_Core with traces of Anat_N

of Avar_Asia_Core1, whereas the other 9 samples were re-grouped as Avar_Asia_Core2 (Figure 2).

According to outgroup f3-statistics, both Avar_Asia_Core groups had highest shared drift with genomes with predominantly North-East Asian (ANA) ancestry (Data S6B), like earlyXiongnu_rest, Ulaanzukh, and Slab Grave.²² It is notable that from the populations with top 50 f3 values, 41 are shared with Hun_Asia_Core; moreover, Avar_Asia_Core1 is in the top 50 populations for both Hun_Asia_Core samples, signifying common deep ancestry of European Huns and Avars.

According to distal qpAdm models, Avar_Asia_Core formed a clade with the Fofonovo_EN and centralMongolia_preBA genomes (Figure 3A; Data S8A), both of which had been modeled from 83%–87% ANA and 12%–17% ANE.²² All data consistently show that Avar_Asia_Core preserved very ancient Mongolian pre-Bronze Age genomes, with ca 90% ANA ancestry.

Most proximodistal qpAdm models (defined in STAR Methods) retained distal sources, as Avar_Asia_Core1 was modeled from 95% UstBelaya_N²⁹ plus 5% Steppe Iron Age (Steppe_IA) and Avar_Asia_Core2 from 80%–92% UstBelaya_N plus 8%–20% Steppe_IA (Data S8B). The exceptional proximal model for Avar_Asia_Core1 indicated Yana_Medieval²⁹ plus Ulaanzukh, whereas for Avar_Asia_Core2, Xianbei_Hun_Berel²¹ plus Kazakhstan_Nomad_Hun_Sarmatian²⁰ ancestries (Figure 3A).

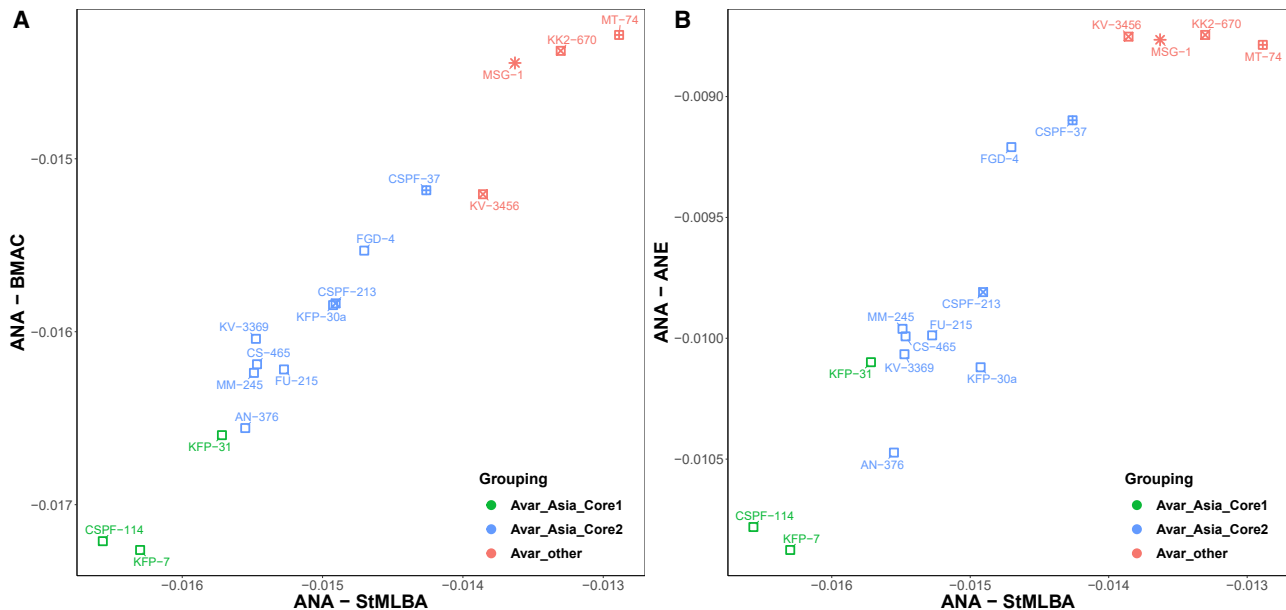


Figure 4. Two-dimensional f_4 -statistics of Avar_Asia_Core individuals

(A) f_4 values from the statistics $f_4(\text{Ethiopia_4500BP, Test; Ulaanzuukh_SlabGrave, MLBA_Sintashta})$ versus $f_4(\text{Ethiopia_4500BP, Test; Ulaanzuukh_SlabGrave, Uzbekistan_BA_Bustan})$ measuring the relative affinities of Test individuals with Steppe_MLBA and BMAC.

(B) f_4 values from the statistics $f_4(\text{Ethiopia_4500BP, Test; Ulaanzuukh_SlabGrave, MLBA_Sintashta})$ versus $f_4(\text{Ethiopia_4500BP, Test; Ulaanzuukh_SlabGrave, Kazakhstan_Eneolithic_Botai})$ measuring the relative affinities of test individuals with Steppe_MLBA and ANE.

The latter model also points to shared ancestries between Huns and Avars.

From the 76 samples in the Avar-cline, 26 could be modeled as a simple 2-way admixture of Avar_Asia_Core and Eur_Core (Figure 3B; Data S8C), indicating that these were admixed descendants of locals and immigrants, whereas further 9 samples required additional Hun- and/or Iranian-related sources. In the remaining 41 models, Hun_Asia_Core and/or Xiongnu sources replaced Avar_Asia_Core (Figure 3B; Data S8D; summarized in Data S1C). Scythian-related sources with significant Iranian ancestries, like Alan, Tian Shan Hun, Tian Shan Saka,²⁰ or Anapa (this study), were ubiquitous in the Avar-cline, but given their low proportion, qpAdm was unable to identify the exact source.

Xiongnu/Hun-related ancestries were more common in certain cemeteries, for example, it was detected in most samples from Hortobágy-Árkus (ARK), Szegvár-Oromdűlő (SZOD), Makó-Mikócsa-halom (MM), and Szarvas-Grexa (SZRV) (Data S8D).

The Conquerors had Ugric, Sarmatian, and Hun ancestries

The Conquest period samples also form a characteristic genetic “Conq-cline” on PCA (Figure 2A). It is positioned north of the Avar-cline, although only reaching the midpoint of the PC1 axis. PC50 clustering identified a single genetic cluster at the Asian extreme of the cline (Data S3) with 12 samples, derived from 9 different cemeteries, that we termed Conq_Asia_Core. This genetic group consists of 6 males and 6 females, and 11 of the 12 individuals belonged to the Conqueror elite according to archaeological evaluation.

The PCA position of Conq_Asia_Core corresponds to modern Bashkirs and Volga Tatars (Figure 2A), and they cluster together

with a wide range of eastern Scythians, western Xiongnu, and Tian Shan Huns,²⁰ which is also supported by ADMIXTURE (Figure 2B).

Two-dimensional f_4 -statistics detected slight genetic differences between Conq_Asia_Core individuals (Figure 5), obtained via multiple gene flow events, as they had different affinity related to Miao (a modern Chinese group) and Ulaanzuukh_SlabGrave (ANA).²² Individuals were arranged linearly along the Miao-ANA cline, suggesting that these ancestries covary in the Conqueror group and thus could have arrived together, most likely from present day Mongolia. As four individuals with highest Miao and ANA affinities also had shifted PCA locations, we set these apart under the name of Conq_Asia_Core2, whereas the rest were re-grouped as Conq_Asia_Core1 (Figure 2). Along the ANE and BMAC axes, the samples showed a more scattered arrangement, although Conq_Asia_Core2 individuals showed somewhat higher BMAC ancestry.

Admixture f_3 -statistics indicated that the main admixture sources of Conq_Asia_Core1 were ancient European populations and ancestors of modern Nganasans (Data S6C). The most likely direct source of the European genomes could be Steppe_MLBA populations, as these distributed European ancestry throughout of the Steppe.³⁰

Outgroup f_3 -statistics revealed that Conq_Asia_Core1 shared highest drift with modern Siberian populations speaking Uralic languages, Nganasan (Samoyedic), Mansi (Ugric), Selkup (Samoyedic), and Enets (Samoyedic) (Data S6E), implicating that Conq_Asia_Core shared evolutionary past with language relatives of modern Hungarians. We also performed f_4 -statistics to test whether the shared evolutionary past was restricted to language relatives. The f_4 -statistics showed that Conq_Asia_Core1 indeed

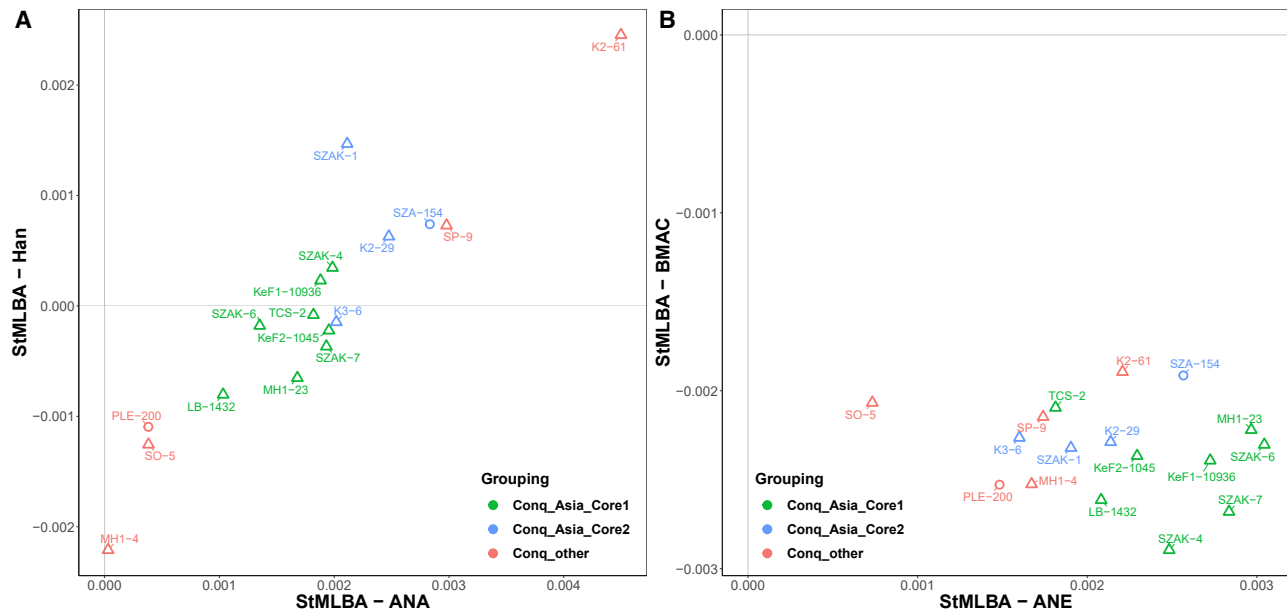


Figure 5. Two-dimensional f_4 -statistics of Conq_Asia_Core individuals

(A) f_4 values from the statistics $f_4(\text{Ethiopia}_{4500\text{BP}}, \text{Test}; \text{MLBA}_{\text{Sintashta}}, \text{Ulaanzuukh}_{\text{SlabGrave}})$ versus $f_4(\text{Ethiopia}_{4500\text{BP}}, \text{Test}; \text{MLBA}_{\text{Sintashta}}, \text{Miao-modern})$ measuring the relative affinities of Test individuals with ANA and Han.

(B) f_4 values from the statistics $f_4(\text{Ethiopia}_{4500\text{BP}}, \text{Test}; \text{Sintashta}, \text{Kazakhstan}_{\text{Eneolithic_Botai}})$ versus $f_4(\text{Ethiopia}_{4500\text{BP}}, \text{Test}; \text{Sintashta}, \text{Uzbekistan}_{\text{BA_Bustan}})$ measuring the relative affinities of Test individuals with ANE and BMAC.

had highest affinity to Mansis, the closest language relatives of Hungarians, but its affinity to Samoyedic-speaking groups was comparable with that of Yeniseian-speaking Kets and Chukotko-Kamchatkan-speaking Koryaks (Data S6G). For this reason, we coanalyzed Mansis with Conq_Asia_Core.

From pre-Iron Age sources, Mansis could be qpAdm modeled from Mezhovskaya,¹⁸ Nganasan, and Botai,²⁸ and Conq_Asia_Core1 from Mezhovskaya, Nganasan, Altai_MLBA_o,²¹ and Mongolia_LBA_CenterWest_4D²³ (Figure 3A; Data S9A and S9B), confirming shared late Bronze Age ancestries of these groups but also signifying that the Nganasan-like ancestry was largely replaced in Conq_Asia_Core by a Scytho-Siberian-like ancestry including BMAC^{21,23} derived from the Altai-Mongolia region. The same analysis did not give passing models for Kets and Koryaks, confirming that they had different genome histories.

From proximal sources, Conq_Asia_Core1 could be consistently modeled from 50% Mansi, 35% Early/Late Sarmatian, and 15% Scytho-Siberian-outlier/Xiongnu/Hun ancestries, and Conq_Asia_Core2 had comparable models with shifted proportions (Figure 3A; Data S9C). As the source populations in these models defined inconsistent time periods, we performed DATES analysis³⁰ to clarify admixture time.

DATES revealed that the Mansi-Sarmatian admixture happened ~53 generations before death of the Conqueror individuals, around 643–431 BCE, apparently corresponding to the Sauromatian/early Sarmatian period. The Mansi-Scythian/Hun-related admixture was dated ~24 generations before death, or 217–315 CE, consistent with the post-Xiongnu, pre-Hun period rather than the Iron Age (Figure 6).

Most individuals of the Conqueror cline proved to be admixed descendants of the immigrants and locals: from the 42 samples

in the Conq-cline, 31 could be modeled as two-way admixtures of Conq_Asia_Core and Eur_Core (Figure 3B; Data S9D; summarized in Data S1C). The remaining samples mostly belonged to the elite, many projecting with the Avar-cline (Figure 2A); of these 5 samples could be modeled from Conq_Asia_Core requiring Hun- and Iranian-associated additional sources. In total, 17 outlier individuals lacked Conq_Asia_Core ancestry, which was replaced with Avar_Asia_Core or Xiongnu/Hun-related sources, accompanied by Iranian-associated 3rd sources (Figure 3B; Data S9E). It seems from our data that Conqueror elite individuals with Hun-related genomes were clustered in certain cemeteries; for example, each sample from Szeged-Óthalom (SEO), Algyó 258-as kútkörzet (AGY), Nagykörös-Fekete-dűlő (NK), Sándorfalva-Eperjes (SE), and Sárrétudvari-Poroshalom (SP) had this ancestry.

Neolithic and Caucasus samples

We have sequenced three new Neolithic genomes from Hungary from three different cemeteries. Two individuals represented the Hungary_Tisza Neolithic culture and one individual the Hungary_Starcevo early Neolithic culture (Data S1). Other genomes had been published previously from each site,³¹ and our genomes have the same ADMIXTURE profile (Data S5) and PCA location (Figure 2) as the previously published samples and also cluster together with Anatolian and European farmers on our PC50 clustering (Data S3).

We have also sequenced 3 samples derived from the Caucasus region (Figure 1) with possible archaeological affinity to the Conquerors. On PCA, Anapa individuals project onto modern samples from the Caucasus (Figure 2), and they cluster together with ancient samples from the Caucasus and BMAC regions

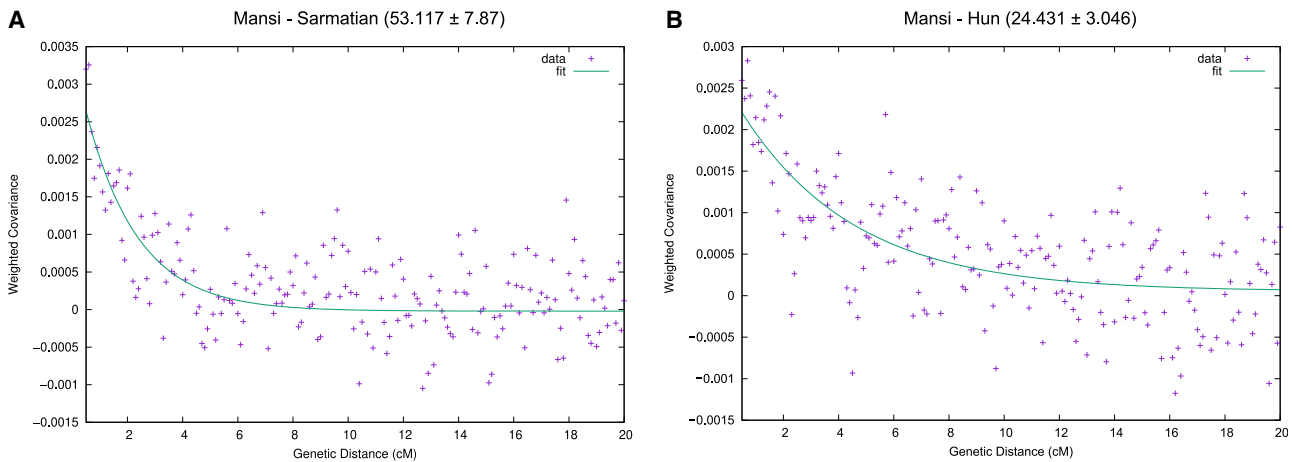


Figure 6. DATES analysis to estimate admixture time

Figures show the weighted ancestry covariant decays for the indicated two-way admixture sources. Curves show the fitted exponential functions, from which the number of generations since admixture are calculated by the program. (A) Mansi-Sarmatian and (B) Mansi-Hun admixture time in generations, in the Conq_Asia_Core target.

(Data S3). The ADMIXTURE profile of the Anapa individuals indicated 33% Iranian, 32% EU_N, 19% ANE, 8% Anat_N, 6% Han 3% Nngasan, and no WHG components (Data S5). These data implied European, Iranian, and Steppe admixtures, the latter including ANA ancestry.

qpAdm revealed that the Anapa individuals carried ancient genomes derived from the Caucasus region, as Armenia_MBA was the majority (76%–96%) source in each model (Data S7F and S7G). In a single lower p value model, Russia_SaltovoMayaki²⁰ formed a clade with our Anapa samples, but this has to be interpreted with caution, as the three available Saltovo-Mayaki genomes have very low coverage (0.029 \times , 0.04 \times , and 0.072 \times). The close proximity to Saltovo-Mayaki samples on PCA as well as their clustering together suggest that proximal sources of Anapa could be similar to that of Saltovo-Mayaki, as these were contemporary individuals and both could be part of the Khazar Khaganate.

Y-chromosome and mtDNA results

The distribution of uniparental markers along the PCA genetic clines shows a general pattern: at the Asian side of the cline, we find individuals with Asian haplogroups (Hgs), whereas at the European side, individuals carry European Hgs. Along the cline from Asia toward Europe, the same trend prevails, decreasing frequency of Asian and increasing of European Hgs. The few exceptions from this rule are nearly always detected in admixed individuals (Data S1A); nevertheless, several individuals in the Eur-cline carried Asian Hgs, testifying distant Asian forefathers. This is especially prominent in the Jánoshida-Tótképuszta (JHT) Late Avar graveyard, where all three males carried R1a1a1b2a (R1a-Z94) Asian Y-Hg, in spite of their European genomes. The Middle Avar MT-17 individual from Madaras-Téglavető in the Eur-cline also carried R1a-Z94, although in this cemetery, all three other males carried N1a1a1a3a with Asian genomes.

Both Hun_Asia_Core individuals (VZ-12673 and MSG-1) carry R1a-Z94 as well as ASZK-1 in the Hun-cline. The

other two published genomes united in Hun_Asia_Core, Kuraly_Hun_380CE²¹ and Kazakhstan_OutTianShanHun,²⁰ carry Hgs R1a-Z94 and Q, respectively, suggesting that these Hgs could be common among the Huns. Considering all published post-Xiongnu Hun era genomes (Hun period nomad, Hun-Sarmatian, Tian Shan Hun,²⁰ and Xianbei-Hun Berel²¹), we counted 10/23 R1a-Z93 and 9/23 Q Hgs, supporting the above observation. These Y-Hgs were most likely inherited from Xiongnus, as these Hgs were frequent among them^{22,32} but were rare in Europe before the Hun period. The rest of our Hun period samples with European genomes carried derivatives of R1a1a1b1, an Hg typical in North-Western Europe, in line with the Germanic affinity of many of these samples shown above.

From the 9 Avar_Asia_Core males, 7 carried the N1a1a1a1a3a (N1a-F4205) Y-Hg, one C2a1a1b1b, and one R1a1a1b~ (very likely R1a-Z94). This confirms that N1a-F4205, most prevalent in modern Chukchis and Buryats,³³ was also prevailing among the Avar elite as shown before.^{34,35} This Hg was also common in members of the Avar-cline and seems to cluster in certain cemeteries. In the Ároktő (ACG), Felgyő (FU), SZOD, Csepel-Kavicsbánya (CS), Kiskörös-Vágóhídi dűlő (KV), Kunpeszér-Felsőpezsér (KFP), Csólyospálos-Felsőpálos (CSPF), Kiskundorozsma-Kettőshatár II (KK2), Tatárszentgyörgy (TTSZ), and Madaras-Téglavető (MT) Avar graveyards, all or the majority of males carried the N1a-F4205 Hg, mostly accompanied with Asian maternal lineages. These cemeteries must have belonged to the immigrant Avar population, whereas the local population seems to have separated, as many Avar period cemeteries show no sign of Asian ancestry. The latter include Mélykút-Sáncdűlő (MS), Szeged-Fehértó A (SZF), Szeged-Kundomb (SZK), Szeged-Makkoserdő (SZM), Kiskundorozsma-Kettőshatár I (KK1), Kiskundorozsma-Daruhalom (KDA), Orosháza-Bónum Téglagyár (OBT), Székkutas-Kápolnadűlő (SZKT), Homokmégy-Halom (HH), Alattyan-Tulát (ALT), Kiskörös-Pohibuj Mackó dűlő (KPM), and Sükösd-Ságod (SSD), in which Asian lineages barely occur. In the SZK, ALT, KK1, OBT, SZKT, HH, and SZM cemeteries, most males belonged to the

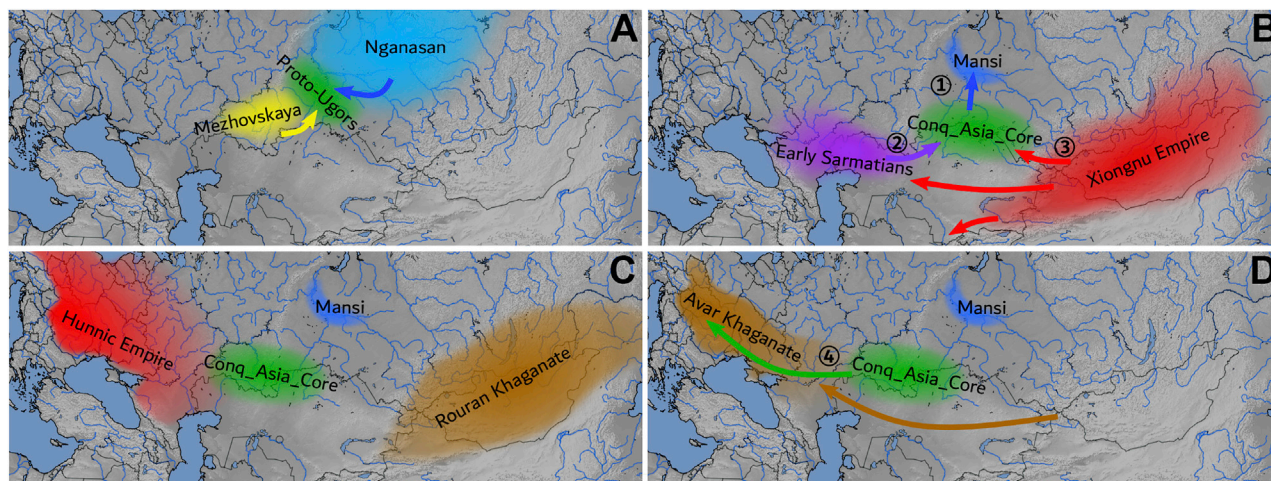


Figure 7. Summary map

(A) Proto-Ugric peoples emerged from the admixture of Mezhovskaya and Nganasan populations in the late Bronze Age.

(B) (1) During the Iron Age Mansis separated and (2) Proto-Conquerors admixed with Early Sarmatians ca 643–431 BCE and (3) with pre-Huns ca 217–315 CE.

(C) By the 5th century, the Xiongnu-derived Hun Empire occupied Eastern Europe, incorporating its population, and the Rouran Khaganate emerged on the former Xiongnu territory.

(D) By the middle 6th century, the Avar Khaganate occupied the territory of the former Hun Empire, incorporating its populations. (4) By the 10th century, Conquerors associated with the remnants of both empires during their migration and within the Carpathian Basin.

E1b1b1a1b1 (E-V13) Hg, which is most prevalent in the Balkan,³⁶ and accordingly, many of the samples from these cemeteries fell in Eur_Core1 or its vicinity, with typical southern European genomes.

There is a third group of Avar period cemeteries representing immigrants from Asia, but with a different genetic background. In males from MM, Dunavecse-Kovacsos dűlő (DK), Árkus Homokbánya (ARK), and SZRV Y-Hgs R1a-Z94 and Q1a2a1 dominated, which seem typical in European Huns, and were mostly accompanied by Asian maternal lineages. These Avar period people could have represented Hun remnants that joined the Avars but isolated in separate communities. These inferences are perfectly in line with genomic data, as most qpAdm models from these cemeteries indicated the presence of Hun_Asia_Core or Xiongnu ancestries (Data S1C). As mentioned above, Hun ancestry was also present in several other cemeteries, just like Hg R1a-Z94, but in those cemeteries, the population was genetically less uniform.

The Conqueror population had a more heterogeneous Hg composition compared with the Huns and Avars. In the 6 Conq_Asia_Core males, we detected three N1a1a1a1a2~, one D1a1a1a1b, one C2a1a1b1b, and one Q1a1a1 Y-Hg, generally accompanied by Asian maternal lineages. Two other N1a1a1a1a2a1c~ Y-Hgs were detected in the SO-5 Conqueror elite and the PLE-95 commoner individuals; thus, this Hg seems specific for the Conqueror group. Obviously, this Hg links Conquerors with Mansis, as had been shown before.³⁷ Another related Y-Hg, N1a1a1a1a4 (M2128), was detected in two Conqueror elite samples from present study as well as from another two Conqueror elite samples in our previous study.³⁴ This Hg is typical for modern Yakuts and occurs with lower frequency among Khantys, Mansis, and Kazakhs,³³ and thus may also link Conquerors with Mansis, although it was also present in one Middle Avar individual. It is notable that

the European Y-Hg I2a1a2b1a1a was also specific for the Conqueror group, especially for the elite as also shown before,³⁴ very often accompanied by Asian maternal lineages, indicating that I2a1a2b1a1a could be more typical for the immigrants than to the local population. Additionally, two other Y-Hgs appeared with notable frequencies among the Conquerors: R1a-Z94 was present in 3 elite and 2 commoner individuals, whereas Hg Q was carried by 3 elite individuals, which may be sign of Hun relations, also detected at the genome level. This result is again in line with genome data, as nearly all Conquest period males with R1a-Z94 or Q Hgs carried Hun-related ancestry.

DISCUSSION

The genomic history of Huns, Avars, and Conquerors revealed in this study is compatible with historical, archaeological, anthropological, and linguistic sources (summarized in Figure 7). Our data show that at least part of the military and social leader strata of both European Huns and Avars likely originated from the area of the former Xiongnu Empire, from present day Mongolia, and both groups can be traced back to early Xiongnu ancestors. Northern Xiongnus were expelled from Mongolia in the second century CE, and during their westward migration, Sarmatians were one of the largest groups they confronted. Sergey Botalov presumed the formation of a Hun-Sarmatian mixed culture in the Ural region before the appearance of Huns in Europe,³⁸ which fits the significant Sarmatian ancestry detected in our Hun samples, although this ancestry had been present in late Xiongnus as well.²² Thus our data are in accordance with the Xiongnu ancestry of European Huns, claimed by several historians.^{39,40} We also detected Goth- or other Germanic-type genomes²⁵ among our Hun period samples, again consistent with historical sources.³⁹

Most of our Avar_Asia_Core individuals represented the early Avar period and half of the “elite” samples belonged to Avar_Asia_Core (Data S1). The other elite samples also contained a high proportion of this ancestry, suggesting that this ancestry could be prevalent among the elite, although also present in common people. The elite preserved very ancient east Asian genomes with well-defined origin, as had been also inferred from Y-Hg data.^{34,35} Our data are compatible with the Rouran origin of the Avar elite,⁵ although the single low-coverage Rouran genome⁴¹ provided a poor fit in the qpAdm models (Data S8B). However, less than half of the Avar-cline individuals had Avar_Asia_Core ancestry, indicating the diverse origin of the Avar population. Our models indicate that the Avars incorporated groups with Xiongnu/Hun_Asia_Core and Iranian-related ancestries, presumably the remnants of the European Huns and Alans or other Iranian peoples on the Pontic Steppe, as suggested by Kim.³⁹ People with different genetic ancestries were seemingly distinguished, as samples with Hun-related genomes were buried in separate cemeteries.

The Conquerors, who arrived in the Carpathian Basin after the Avars, had a distinct genomic background with elevated levels of western Eurasian admixture. Their core population carried very similar genomes to modern Bashkirs and Tatars, in agreement with our previous results from uniparental markers.^{34,42} Their genomes were shaped by several admixture events, of which the most fundamental was the Mezhovskaya-Nganasan admixture around the late Bronze Age, leading to the formation of a “proto-Ugric” gene pool. This was part of a general demographic process, when most Steppe_MLBA populations received an eastern Khovsgol-related Siberian influx together with a BMAC influx,²¹ and ANA-related admixture became ubiquitous on the eastern Steppe,³⁰ establishing the Scytho-Siberian gene pool. Consequently proto-Ugric groups could be part of the early Scytho-Siberian societies of the late Bronze Age to early Iron Age steppe-forest zone in the northern Kazakhstan region, in the proximity of the Mezhovskaya territory.

Our data support linguistic models, which predicted that Conquerors and Mansis had a common early history.^{7,43} Then Mansis migrated northward, probably during the Iron Age, and in isolation, they preserved their Bronze-Age genomes. In contrast, the Conquerors stayed at the steppe-forest zone and admixed with Iranian-speaking early Sarmatians, also attested by the presence of Iranian loanwords in the Hungarian language.⁴³ This admixture likely happened when Sarmatians rose to power and started to integrate their neighboring tribes before they occupied the Pontic-Caspian Steppe.⁴⁴

All analysis consistently indicated that the ancestors of Conquerors further admixed with a group from Mongolia, carrying Han-ANA-related ancestry, which could be identified with ancestors of European Huns. This admixture likely happened before the Huns arrived in the Volga region (370 CE) and integrated local tribes east of the Urals, including Sarmatians and the ancestors of Conquerors. These data are compatible with a Conqueror homeland around the Ural region, in the vicinity of early Sarmatians, along the migration route of the Huns, as had been surmised from the phylogenetic connections between the Conquerors and individuals of the Kushnarenkovo-Karayakupovo culture in the Trans-Uralic Uyelgi cemetery.⁴⁵ Recently, a Nganasan-like shared Siberian genetic ancestry was detected

in all Uralic-speaking populations, Hungarians being an exception.⁴⁶ Our data resolve this paradox by showing that the core population of conquering Hungarians had high Nganasan ancestry. The fact that this is negligible in modern Hungarians is likely due to the substantially smaller number of immigrants compared with the local population.

The large number of genetic outliers with Hun_Asia_Core ancestry in both Avars and Conquerors testifies that these successive nomadic groups were indeed assembled from overlapping populations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Ancient samples
- METHOD DETAILS
 - Accelerator mass spectrometry radiocarbon dating
 - Ancient DNA laboratory work
 - NGS library construction
 - DNA sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Bioinformatical processing
 - Quality assessment of ancient sequences
 - Uniparental haplogroup assignment
 - Genetic sex determination
 - Estimation of genetic relatedness
 - Population genetic analysis
 - Principal Component Analysis (PCA)
 - Unsupervised Admixture
 - Hierarchical Ward clustering
 - Admixture modeling using qpAdm
 - f3-statistics
 - Two dimensional f4-statistics
 - Dating admixture time with DATES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.04.093>.

ACKNOWLEDGMENTS

We are grateful to our archaeologist colleagues Gabriella M. Lezsák and Andrej Novicsihin for providing us with the Anapa samples, Gábor Lőrinczy for his help regarding the Avar material, and Zsófia Rácz for her help with the Hun period samples. We are thankful to all the museum curators and archaeologists who provided bone material for this study: Herman Ottó Museum Miskolc, Laczkó Dezső Museum Veszprém, Budapest History Museum, Ferenczy Museum Szentendre, Dobó István Castle Museum Eger, Jóna András Museum Nyíregyháza, Katona József Museum Kecskemét, and Janus Pannonius Museum Pécs. This research was funded by grants from the National Research, Development and Innovation Office (K-124350 to T.T. and TUDFO/5157-1/2019-ITM and TKP2020-NKA-23 to E.N.), The House of Árpád

Programme (2018–2023) Scientific Subproject: V.1. Anthropological-Genetic portrayal of Hungarians in the Árpadian Age to T.T. and no. VI/1878/2020. certificate number grants to E.N.

AUTHOR CONTRIBUTIONS

Conceptualization, supervision, project administration, and funding acquisition, T.T. and E. Neparáczki; data curation and software, Z.M., T.K., and E. Nyerki; formal analysis, validation, and methodology, Z.M., O.S., and T.T.; resources; Z.G., C.H., S.V., L.K., C.B., C.S., G.S., E.G., A.P.K., B.G., B.N.K., S.S.G., P.T., B.T., A.M., G.P., Z.B., A.Z., and F.M.; investigation, K.M., G.I.B.V., B.K., E. Neparáczki, O.S., P.L.N., I.N., D.L., A.G., and R.G.; visualization, Z.M. and O.S.; writing – original draft, T.T. with considerable input from I.R.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

P.L.N. from Praxis Genomics LLC, I.N. and D.L. from SeqOmics Biotechnology Ltd., and Z.G. from Ásatárs Ltd. were not directly involved in the design of the experiments, data analysis, and evaluation. These affiliations do not alter our adherence to *Current Biology's* policies on sharing data and materials.

Received: January 28, 2022

Revised: March 10, 2022

Accepted: April 28, 2022

Published: May 25, 2022

REFERENCES

- Schmauder, M. (2015). Huns, Avars, Hungarians – reflections on the interaction between steppe empires in southeast Europe and the late roman to early byzantine empires. In *Complexity of Interaction along the Eurasian Steppe Zone in the First C.E. Millennium*, J. Bemmann, and M. Schmauder, eds. (Bonn Contributions to Asian Archaeology), pp. 671–692.
- Szentpétery, I. (1937). *Scriptores Rerum Hungaricarum Tempore Ducum Regnumque Stirpis Arpadianae Gestarum* (Acad. Litter. Hungarica).
- Hóman, B. (1925). *A magyar hún-hagyomány és hún monda* (Studium).
- De La Vaissière, É. (2014). *The Steppe World and the Rise of the Huns* (Cambridge University Press), pp. 175–192.
- Golden, P.B. (2013). Some notes on the Avars and Rouran (Editura Universitatii “Alexandru Ioan Cuza”). In *The Steppe Lands and the World Beyond Them: Studies in Honor of Victor Spinei on His 70th Birthday*, F. Curta, and B.-P. Maleon, eds. (Editura Universitatii “Alexandru Ioan Cuza”), pp. 43–66.
- Golden, P.B. (1992). *An Introduction to the History of the Turkic Peoples: Ethnogenesis and State Formation in Medieval and Early Modern Eurasia and the Middle East* (O. Harrassowitz).
- Róna-Tas, A. (1999). *Hungarians and Europe in the Early Middle Ages: An Introduction to Early Hungarian History* (Central European University Press).
- Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjänen, K., and Wahlberg, N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *J. Evol. Biol.* 26, 1244–1253.
- Fodor, I. (2010). A hún-magyar rokonság elmélete. In *Egyezünk ki a múlttal! Műhelybeszélgetések történelmi mítoszainkról, tévhiteinkről*, L. Lőrinc, ed. (Történelemtanárok Egylete), pp. 165–168.
- Rady, M. (2018). Attila and the hun tradition in Hungarian medieval texts. In *Project MUSE - Studies on the Illuminated Chronicle*, J.M. Bak, and L. Veszprémy, eds. (Central European Medieval Texts), pp. 127–138.
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684.
- Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16, 224.
- Korneliusson, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 15, 1–13.
- Nyerki, E., Kalmár, T., Schütz, O., Lima, R.M., Neparáczki, E., Török, T., and Maróti, Z. (2022). An optimized method to infer relatedness up to the 5th degree from low coverage ancient human genomes. Preprint at bioRxiv. <https://doi.org/10.1101/2022.02.11.480116>.
- Amorim, C.E.G., Vai, S., Posth, C., Modi, A., Koncz, I., Hakenbeck, S., La Rocca, M.C., Mende, B., Bobo, D., Pohl, W., et al. (2018). Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* 9, 1–11.
- Antonio, M.L., Gao, Z., Moots, H.M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R.C., et al. (2019). Ancient Rome: a genetic crossroads of Europe and the Mediterranean. *Science* 366, 708–714.
- Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and Mycenaeans. *Nature* 548, 214–218.
- Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172.
- Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190–196.
- Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliusson, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374.
- Gneccchi-Ruscione, G.A., Khussainova, E., Kahbatkyzy, N., Musralina, L., Spyrou, M.A., Bianco, R.A., Radzeviciute, R., Martins, N.F.G., Freund, C., Iksan, O., et al. (2021). Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Sci. Adv.* 7, 4414–4440.
- Jeong, C., Wang, K., Wilkin, S., Taylor, W.T.T., Miller, B.K., Bemmann, J.H., Stahl, R., Chiovelli, C., Knolle, F., Ulziibayar, S., et al. (2020). A dynamic 6,000-year genetic history of Eurasia's eastern steppe. *Cell* 183, 890–904.e29.
- Wang, C.C., Yeh, H.Y., Popov, A.N., Zhang, H.Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M., et al. (2021). Genomic insights into the formation of human populations in East Asia. *Nature* 597, 413–419.
- Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., Hudson, M.J., et al. (2020). Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun.* 11, 1–9.
- Järve, M., Saag, L., Scheib, C.L., Pathak, A.K., Montinaro, F., Pagani, L., Flores, R., Guellil, M., Saag, L., Tambets, K., et al. (2019). 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.e10.
- Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., et al. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* 7, 1–9.
- Balogh, C. (2019). Új szempont a kora Avar hatalmi központ továbbélésének kérdéséhez – az Avar fegyveres réteg temetései a Duna-Tisza közén – new light on the possible survival of the early Avar power centre – burials of Avar warriors in the Danube-Tisza interfluvium. In *Hatalmi Központok Az Avar Kaganátusban – Power Centres of the Avar Khaganate*, C. Balogh, J. Szentpétery, and E. Wicker, eds. (Katona József Múzeum Kecskemét), pp. 115–138.
- De Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamos, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360, eaar7711.

29. Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G., et al. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature* 570, 182–188.
30. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. *Science* 365, eaat7487.
31. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551, 368–372.
32. Keyser, C., Zvenigorosky, V., Gonzalez, A., Fausser, J.L., Jagorel, F., Gérard, P., Tsagaan, T., Duchesne, S., Crubézy, E., and Ludes, B. (2021). Genetic evidence suggests a sense of family, parity and conquest in the Xiongnu Iron Age nomads of Mongolia. *Hum. Genet.* 140, 349–359.
33. Ilumäe, A.M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., Saag, L., Agdzhoyan, A., Kushniarevich, A., Litvinov, S., et al. (2016). Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. *Am. J. Hum. Genet.* 99, 163–173.
34. Neparáczki, E., Maróti, Z., Kalmár, T., Maár, K., Nagy, I., Latinovics, D., Kustár, Á., Pálfi, G., Molnár, E., Marcsik, A., et al. (2019). Y-chromosome haplogroups from Hun, Avar and conquering Hungarian period nomadic people of the Carpathian Basin. *Sci. Rep.* 9, 16569.
35. Csáky, V., Gerber, D., Koncz, I., Csiky, G., Mende, B.G., Szeifert, B., Egyed, B., Pamjav, H., Marcsik, A., Molnár, E., et al. (2020). Genetic insights into the social organisation of the Avar period elite in the 7th century AD Carpathian Basin. *Sci. Rep.* 10, 1–14.
36. Cruciani, F., La Fratta, R., Trombetta, B., Santolamazza, P., Sellitto, D., Colomb, E.B., Dugoujon, J.M., Crivellaro, F., Benincasa, T., Pascone, R., et al. (2007). Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol. Biol. Evol.* 24, 1300–1311.
37. Post, H., Németh, E., Klima, L., Flores, R., Fehér, T., Türk, A., Székely, G., Sahakyan, H., Mondal, M., Montinaro, F., et al. (2019). Y-chromosomal connection between Hungarians and geographically distant populations of the Ural Mountain region and West Siberia. *Sci. Rep.* 9, 1–10.
38. Botalov, S.G., and Gutsalov, S. (2000). Hunno-Sarmatians of the Ural-Kazakh Steppes (Прифεί).
39. Kim, H.J. (2013). *The Huns, Rome and the Birth of Europe* (Cambridge University Press).
40. De La Vaissière, É. (2014). The steppe world and the rise of the Huns. In *Cambridge Companion to Age Attila*, M. Maas, ed. (Cambridge University Press), pp. 175–192.
41. Li, J., Zhang, Y., Zhao, Y., Chen, Y., Ochir, A., Sarenbilge, Z., Zhu, H., and Zhou, H. (2018). The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population. *Am. J. Phys. Anthropol.* 166, 895–905.
42. Neparáczki, E., Maróti, Z., Kalmár, T., Kocsy, K., Maár, K., Bihari, P., Nagy, I., Fóthi, E., Pap, I., Kustár, Á., et al. (2018). Mitogenomic data indicate admixture components of Central-Inner Asian and Srubnaya origin in the conquering Hungarians. *PLoS One* 13, e0205920.
43. Abondolo, D.M. (1998). *The Uralic Languages* (Routledge).
44. Istvánovits, E., and Kulcsár, V. (2017). *Sarmatians: History and Archaeology of a Forgotten People* (Römisch-Germanisches Zentralmuseum).
45. Csáky, V., Gerber, D., Szeifert, B., Egyed, B., Stégmár, B., Botalov, S.G., Grudochko, I.V., Matveeva, N.P., Zelenkov, A.S., Sleptsova, A.V., et al. (2020). Early medieval genetic data from Ural region evaluated in the light of archaeological evidence of ancient Hungarians. *Sci. Rep.* 10, 1–14.
46. Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A.M., Rootsi, S., Honkola, T., Vesakoski, O., Atkinson, Q., Skoglund, P., Kushniarevich, A., et al. (2018). Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biol.* 19, 139.
47. Reich Lab, David (2020). *Allen ancient DNA resource*. <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>.
48. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12.
49. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
50. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
51. Broad Institute (2016). *Picard tools*. <https://broadinstitute.github.io/picard/>.
52. Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., and Wegmann, D. (2017). ATLAS: analysis tools for low-depth and ancient samples. Preprint at bioRxiv. <https://doi.org/10.1101/105346>.
53. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44, W58–W63.
54. Ralf, A., Montiel González, D., Zhong, K., and Kayser, M. (2018). Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. *Mol. Biol. Evol.* 35, 1291–1294.
55. Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34, 867–868.
56. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210, 719–731.
57. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
58. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.
59. R Development Core Team. (2015). *R: a language and environment for statistical computing* (R Foundation for Statistical Computing).
60. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.
61. Kuhn, J.M.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PLoS One* 13, e0195491.
62. Molnár, M., Janovics, R., Major, I., Orsovskzi, J., Gönczi, R., Veres, M., Leonard, A.G., Castle, S.M., Lange, T.E., Wacker, L., et al. (2013). Status report of the new AMS 14C sample preparation Lab of the Hertelendi Laboratory of Environmental Studies (Debrecen, Hungary). *Radiocarbon* 55, 665–676.
63. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* 62, 725–757.
64. Maár, K., Varga, G.I.B., Kovács, B., Schütz, O., Maróti, Z., Kalmár, T., Nyerki, E., Nagy, I., Latinovics, D., Tihanyi, B., et al. (2021). Maternal lineages from 10–11th century commoner cemeteries of the Carpathian Basin. *Genes* 12, 460.
65. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, pdb.prot5448.
66. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40, e3.
67. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130624.

68. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* *40*, 4477–4482.
69. Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.C., Ixan, O., Khussainova, E., et al. (2019). The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* *3*, 966–976.
70. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., Roudier, P., Gonzalez, J., and Kozłowski, K. (2018). “Finding Groups in Data”: Cluster Analysis Extended Rousseeuw et al. <https://svn.r-project.org/R-packages/trunk/cluster/>.
71. Unterländer, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanová, Z., Groß, M., Sell, C., Blöcher, J., Kirsanow, K., Rohland, N., et al. (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat. Commun.* *8*, 14615.
72. Krzewińska, M., Kiliņ, G.M., Juras, A., Koptekin, D., Chyleński, M., Nikitin, A.G., Shcherbakov, N., Shuteleva, I., Leonova, T., Kraeva, L., et al. (2018). Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci. Adv.* *4*, eaat4457.
73. Harney, É., Patterson, N., Reich, D., and Wakeley, J. (2021). Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* *217*, iyaa045.
74. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Orlando, L., Metspalu, E., et al. (2014). Upper palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* *505*, 87–91.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human archaeological remains	This paper	N/A
Critical commercial assays		
MinElute PCR Purification Kit	QIAGEN	Cat No./ID: 28006
Accuprime Pfx Supermix	ThermoFisher Scientific	Cat. No: 12344040
Deposited data		
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Modern comparison dataset	Allen Ancient DNA Resource (Version v42.4) ⁴⁷	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data
Ancient comparison dataset	Allen Ancient DNA Resource (Version v42.4) ⁴⁷	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data
Ancient comparison dataset	Gnecchi-Ruscone et al. ²¹	https://www.ebi.ac.uk/ena/browser/view/PRJEB42930
Ancient comparison dataset	Jeong et al. ²²	https://www.ebi.ac.uk/ena/browser/view/PRJEB35748
Ancient comparison dataset	Wang et al. ²³	https://www.ebi.ac.uk/ena/browser/view/PRJEB42781
Ancient comparison dataset	Ning et al. ²⁴	https://www.ebi.ac.uk/ena/browser/view/PRJEB36297
Newly published ancient genomes	This paper	https://www.ebi.ac.uk/ena/browser/view/PRJEB499771
Oligonucleotides		
Illumina specific adapters	Custom synthesized	https://www.sigmaldrich.com/HU/en/product/sigma/oligo?lang=en&region=US&gclid=CjwKCAiAgvKQBhBbEiwAaPQw3FDDFnRPc3WV75qapsXvcTxxzBXy48atqyb6Xi5f_8e6Df2EJlONNhoCmzIQAvD_BwE
Software and algorithms		
Cutadapt	Martin ⁴⁸	https://cutadapt.readthedocs.io/en/stable/#
Burrow-Wheels-Aligner	Li and Durbin ⁴⁹	http://bio-bwa.sourceforge.net/
samtools	Li et al. ⁵⁰	http://www.htslib.org/
PICARD tools	Broad Institute ⁵¹	https://github.com/broadinstitute/picard
ATLAS software package	Link et al. ⁵²	https://bitbucket.org/wegmannlab/atlas/wiki/Home
MapDamage 2.0	Jónsson et al. ¹¹	https://ginolhac.github.io/mapDamage/
Schmutzi software package	Renaud et al. ¹²	https://github.com/grenaud/schmutzi
ANGSD software package	Korneliussen et al. ¹³	https://github.com/ANGSD/angsd
HaploGrep 2	Weissensteiner et al. ⁵³	https://haplogrep.i-med.ac.at/category/haplogrep2/
Yleaf software tool	Ralf et al. ⁵⁴	https://github.com/genid/Yleaf
mosdepth software	Pedersen and Quinlan ⁵⁵	https://github.com/brentp/mosdepth
PCAngsd software	Meisner and Albrechtsen ⁵⁶	https://github.com/Rosemeis/pcangsd
RcppCNPY R package	N/A	https://rdocumentation.org/packages/RcppCNPY/versions/0.2.10
smartpca	Patterson et al. ⁵⁷	https://github.com/chrchang/eigensoft/blob/master/POPGEN/README

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ADMIXTURE software	Alexander et al. ⁵⁸	https://dalexander.github.io/admixture/
R 3.6.3	R core development team ⁵⁹	https://cran.r-project.org/bin/windows/base/old/3.6.3/
ADMIXTOOLS software package	Patterson et al. ⁶⁰	https://github.com/DReichLab/AdmixTools
DATES algorithm	Narasimhan et al. ³⁰	https://github.com/priyamoorjani/DATES/tree/v753
READ algorithm	Kuhn et al. ⁶¹	https://doi.org/10.1371/journal.pone.0195491.s007

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tibor Török (torokt@bio.u-szeged.hu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Aligned sequence data have been deposited at European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) under accession number PRJEB49971 and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ancient samples

We present genome-wide data of 265 ancient individuals from the Migration Period of the Carpathian Basin between the 5th and 11th centuries and 3 individuals from the 9-10th century Caucasus with archaeological affinity to the Conquering Hungarians. We also sequenced 3 Neolithic individuals from the Carpathian Basin. The 265 Migration Period samples represent the following time range: 9 samples from the Hun Period (5th century), 40 from the early Avar Period (7th century), 33 from the middle Avar Period (8th century), 70 from the late Avar Period (8-9th century), 48 from 10th century Conquering Hungarian elite cemeteries, 65 from commoner cemeteries of the Hungarian conquer-early Árpáadian Period (10-11th centuries).

The majority of samples were collected from the Great Hungarian Plain (Alföld), the westernmost extension of the Eurasian steppe, which provided favorable ground for the arriving waves of nomadic groups. Cemeteries and individual samples were chosen on archaeological, anthropological and regional basis. We made an effort to assemble a sample collection from each period representing a) all possible geographical sub-regions, b) all archeological types, c) all anthropological types. From large cemeteries we selected individuals with the same criteria and possibly from all part of the cemetery with the following bias: We preferably choose samples with good bone preservation, archaeologically well described ones (with grave goods) and males (for Y-chromosomal data). Nevertheless, we took care to also include females and samples without grave goods, though these are definitely underrepresented in our collection.

The human bone material used for ancient DNA analysis in this study were obtained from anthropological collections or museums, with the permission of the custodians in each case. In addition, we also contacted the archaeologists who excavated and described the samples, as well as the anthropologists who published anthropological details. In most cases these experts became co-authors of the paper, who provided the archaeological background, which is detailed in [Methods S1](#).

METHOD DETAILS

Accelerator mass spectrometry radiocarbon dating

Here we report 73 radiocarbon dates, of which 50 are first reported in this paper. The sampled bone fragments were measured by accelerator mass spectrometry (AMS) in the AMS laboratory of the Institute for Nuclear Research, Hungarian Academy of Sciences, Debrecen, Hungary. Technical details concerning the sample preparation and measurement are given in Kuhn et al.⁶² Several

radiocarbon measurements were done in the Radiocarbon AMS facility of the Center for Applied Isotope Studies, University of Georgia (n = 6); technical details concerning the sample preparation and measurement are available here: <https://cais.uga.edu/facilities/radiocarbon-ams-facility/>). The conventional radiocarbon data were calibrated with the OxCal 4.4 software (<https://c14.arch.ox.ac.uk/oxcal/OxCal.html>, date of calibration: 4th of August 2021) with IntCal 20 settings.⁶³ Besides, we collected all previously published radiocarbon data related to the samples of our study.

Ancient DNA laboratory work

All pre-PCR steps were carried out in the dedicated ancient DNA facilities of the Department of Genetics, University of Szeged and Department of Archaeogenetics, Institute of Hungarian Research, Hungary. Mitogenome or Y-chromosome data had been published from many of the samples used in this study,^{42,64} and we sequenced whole genomes from the same libraries, whose preparations had been described in the above papers.

For the rest of the samples we used the following modified protocol. DNA was extracted from bone powder collected from petrous bone or tooth cementum. 100 mg bone powder was predigested in 3 ml 0.5 M EDTA 100 µg/ml Proteinase K for 30 minutes at 48°C, to increase the proportion of endogenous DNA. After pelleting, the powder was solubilized for 72 hours at 48°C, in extraction buffer containing 0.45 M EDTA, 250 µg/ml Proteinase K and 1% Triton X-100. Then 12 ml binding buffer was added to the extract, containing 5 M GuHCl, 90 mM NaOAc, 40% isopropanol and 0.05% Tween-20, and DNA was purified on Qiagen MinElute columns.

NGS library construction

Partial UDG treated libraries were prepared as described in Neparáczkiet al.⁴² In short we used the double stranded library protocol of Meyer and Kircher⁶⁵ with double indexing,⁶⁶ except that all purifications were done with MinElute columns. We also applied partial UDG treatment of Rohland et al.,⁶⁷ but decreased the recommended USER and UGI concentrations to half (0.03 U/µL) and at the same time increased the incubation time from 30 to 40 minutes. The reaction was incubated at 37°C for 40 minutes in PCR machine, with 40°C lid temperature.

Then 1,8 µL UGI (Uracil Glycosylase Inhibitor, 2U/µL NEB) was added to the reaction, which was further incubated at 37°C for 40 minutes. Next blunt-end repair was done by adding 3 µL T4 polynucleotide kinase (10 U/µL) and 1,2 µL T4 DNA polymerase-t (5 U/µL) to each reaction followed by incubation in PCR machine at 25°C for 15 minutes, and another incubation at 12°C for 5 minutes and cooling to 4°C. After adding 350 µl MinElute PB buffer (QIAGEN) to the reaction, it was purified on MinElute columns, and the DNA was eluted in 20 µl EB prewarmed to 55°C. Adapter ligation and adapter fill-in was done as in Meyer and Kircher.⁶⁵

The library preamplification step was omitted, and libraries were directly double indexed in one PCR-step after the adapter fill with Accuprime Pfx Supermix, containing 10mg/ml BSA and 200nM indexing P5 and P7 primers, in the following cycles: 95°C 5 minutes, 12 times 95°C 15 sec, 60°C 30 sec and 68°C 3 sec, followed by 5 minute extension at 68°C. The indexed libraries were purified on MinElute columns and eluted in 20µL EB buffer (Qiagen).

DNA sequencing

Quantity measurements of the DNA extracts and libraries were performed with the Qubit fluorometric quantification system. The library fragment distribution was checked on TapeStation 2200 (Agilent). We estimated the endogenous human DNA content of each library with low coverage shotgun sequencing generated on iSeq 100 (Illumina) platform. Whole genome sequencing was performed on HiSeqX or NovaSeq 6000 Systems (Illumina) using paired-end sequencing method (2x150bp) following the manufacturer's recommendations.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatical processing

Sequencing adapters were trimmed with the Cutadapt software⁴⁸ [<https://doi.org/10.14806/ej.17.1.200>] and sequences shorter than 25 nucleotides were removed. The raw reads were aligned to the GRCh37 (hs37d5) reference genome by Burrow-Wheels-Aligner (v 0.7.17),⁴⁹ using the MEM command with reseeding disabled.

The current high throughput aligners were designed to work on modern uncontaminated DNA, thus they will find the best fit for all DNA fragments that partially match the reference genome. Since the human genome is very complex, the majority of the exogenous, non human reads will be aligned to it with partial fits. The mapping quality (MAPQ) is a metric that describes only the alignment quality of the matching part of the sequence, regardless of any overhanging (soft, hard clipped) parts. As a consequence, non human reads will receive high MAPQ values, when the length of the aligned part is above 29-30 bases and the mismatches are not excessive, despite of the overall bad matching due to the overhanging ends. For this reason, MAPQ based filtering of exogenous DNA may lead to excessive bias. To avoid this bias we filtered our alignments, based on the proportion of the full length reads matching the human reference genome. To remove exogenous DNA only the primary alignments with $\geq 90\%$ identity to reference were considered in all downstream analysis. This way we could eliminate most exogenous DNA with partial matching, while still keeping human reads with PMD and real SNP variations. Although this method also excludes reads corresponding to true human structural variants, but these are negligible. This strategy results in much more reliable variant calls compared to the simple MAPQ based filtration in case of ancient DNA.

From paired-end sequencing data we only kept properly paired primary alignments. Sequences from different lanes with their unique read groups were merged by samtools.⁵⁰ PICARD tools⁵¹ were used to mark duplicates. In case of paired-end reads we used the ATLAS software package⁵² mergeReads task with the options “updateQuality mergingMethod=keepRandomRead” to randomly exclude overlapping portions of paired-end reads, to mitigate potential random pseudo haploidization bias.

Quality assessment of ancient sequences

Ancient DNA damage patterns were assessed using MapDamage 2.0¹¹ (Data S1B), and read quality scores were modified with the Rescale option to account for post-mortem damage. Mitochondrial contamination was estimated with the Schmutzi software package.¹² Contamination for the male samples was also assessed by the ANGSD X chromosome contamination method,¹³ with the “-r X:5000000-154900000 -doCounts 1 -iCounts 1 -minMapQ 30 -minQ 20 -setMinDepth 2” options.

Contamination estimations are detailed in Data S1A. The estimated contaminations ranged from 0 to 0.42 with an average of 0.017 in the case of Schmutzi and from 0 to 0.13 with an average of 0.014 in the case of ANGSD. We detected negligible contamination in all but 7 samples. Schmutzi estimated significant contamination in 6 samples, but in 3 male samples of these ANGSD measured low X-contamination. These contradictory estimates may be explained by the UDG overtreatment of these libraries, as Schmutzi reckons Uracil free molecules as derived from contamination. In another male sample ANGSD measured significant contamination, which was not confirmed by Schmutzi.

Uniparental haplogroup assignment

Mitochondrial haplogroup determination was performed with the HaploGrep 2 (version 2.1.25) software,⁵³ using the consensus endogen fasta files resulting from the Schmutzi Bayesian algorithm. The Y haplogroup assessment was performed with the Yleaf software tool,⁵⁴ updated with the ISOGG2020 Y tree dataset.

Genetic sex determination

Biological sex was assessed with the method described in Skoglund et al.⁶⁸ Fragment length of paired-end data and average genome coverages (all, X, Y, mitochondrial) was assessed by the ATLAS software package⁵² using the BAMDiagnostics task. Detailed coverage distribution of autosomal, X, Y, mitochondrial chromosomes was calculated by the mosdepth software.⁵⁵

Estimation of genetic relatedness

Presence of close relatives in the dataset interferes with unsupervised ADMIXTURE and population genetic analysis, therefore we identified close kins and just one of them was left in the dataset (Table S1). We performed kinship analysis using the 1240K dataset and the PCAngsd software (version 0.931)⁵⁶ from the ANGSD package with the “-inbred 1 -kinship” options. We used the R (version 4.1.2); the RcppCNPy R package (version 0.2.10) to import the Numpy output files of PCAngsd. Though the PCAngsd software in the ANGSD package has not been a standard method to infer kinship, nevertheless it outperforms all other methods, as we have shown in our manuscript.¹⁴ Close kinship relations identified with PCAngsd were also confirmed with the READ (Relationship Estimation from Ancient DNA) method⁵¹ (Table S1).

Population genetic analysis

The newly sequenced genomes were merged and co-analyzed with 2364 ancient (Data S3) and 1397 modern Eurasian genomes (Table S2), most of which were downloaded from the Allen Ancient DNA Resource (Version v42.4).⁴⁷ We also downloaded the Human Origins dataset (HO, 600K SNP-s) and/or the 1240K datasets published in Gneccchi-Ruscone et al.,²¹ Jeong et al.,²² and Wang et al.²³ As the HO SNPs are fully contained in the larger 1240K set we filtered out the HO data when only the 1240K dataset was published. In case the Reich data set contained preprint data of the same individual published later, we always used the published genotypes. Since some dataset contained diploid, and mixed call variants we performed random pseudo haploidization of all data prior to downstream analysis. To avoid bias of PE sequencing in random haploid calling, we used the ATLAS software “mergeReads” task with parameters “updateQuality mergingMethod=keepRandom” to set BASE QUALITY to 0 of overlapping portion of a random mate of PE reads. Then used ANGSD (software package version: 0.931-10-g09a0fc5)¹³ for random allele calling with the options “--doHaploCall 1 -doCounts 1”, with the default BAM quality filters (including -minQ 20; Discard bases with base quality below) on these mate pair merged BAMs. Since every DNA fragment is represented by only one copy (with BASE QUALITY > 0), this method is equal to random allele calling.

Most of the analysis was done with the HO dataset, as most modern genomes are confined to this dataset, however, we run some of the f-statistics with the 1240K data if these were available.

Principal Component Analysis (PCA)

We used the modern Eurasian genome data published in Jeong et al.,⁶⁹ confined to the HO dataset, to draw a modern PCA background on which ancient samples could be projected. However, in order to obtain the best separation of our samples in the PC1-PC2 dimensions, South-East Asian and Near Eastern populations were left out, and generally just 10 individuals were selected from each of the remaining populations, leaving 1397 modern individuals from 179 modern populations in the analysis (Table S2).

PCA Eigen vectors were calculated from 1397 pseudo-haploidized modern genomes with smartpca (EIGENSOFT version 7.2.1).⁵⁷ Before projecting pseudo-haploidized ancient genomes, we excluded all relatives, and used the individuals with best genome

coverage. All ancient genomes were projected on the modern background with the “Isqproject: YES and inbreed: YES” options. Since the ancient samples were projected, we used a more relaxed genotyping threshold (>50k genotyped markers) to exclude samples only where the results could be questionable due to the low coverage.

Unsupervised Admixture

We carried out unsupervised admixture with 3277 genomes including 1010 modern and 2027 ancient published genomes plus 240 ones from present study, excluding all published relatives from each dataset. For this analysis we used the autosomal variants of the HO dataset as many relevant modern populations are missing from the 1240K set. We set strict criteria for the selection of individual samples to minimize bias and maximize the information content of our dataset. We excluded all samples with QUESTIONABLE flag or Ignore tag based on the annotation file of the dataset to remove possibly contaminated samples and population outliers. To compose a balanced high quality dataset furthermore we restricted the selection to maximum 10 individuals per populations and excluded all poorly genotyped samples (<150K genotyped markers).

To prepare the final marker set for ADMIXTURE analysis we removed variants with very low frequency (MAF <0.005) leaving 471,625 autosomal variants. We pruned 116,237 variants in linkage disequilibrium using PLINK with the options “*-indep-pairwise 200 10 0.25*” leaving a final 355388 markers for the 3277 individuals. The total genotyping rate of this high quality dataset was 0.811831. We performed the unsupervised ADMIXTURE analysis for K=3-12 in 30 parallel runs with the ADMIXTURE software (version 1.3.0)⁵⁸ and selected the lowest cross validation error model (K=7) with the highest log-likelihood run for visualization.

Hierarchical Ward clustering

As qpAdm is sensitive both to genome components and their proportions present in the source and Test populations, it works best if genetically homogenous populations are used in the analysis. Similar Admixture composition and small PC1-PC2 distances may indicate population relations derived from shared genetic ancestry, but these data do not have enough resolution to identify homogenous genome subsets which can be merged into populations. Though the first two PCA Eigenvalues capture the highest levels of variation in the data, in our analysis subsequent Eigenvalues had comparable magnitude to the second one, indicating that lower Eigenvectors still harbored significant additional genetic information, which must be taken into account. Individuals with most similar genomes should be closest in the entire PC space, which can be measured mathematically with the combined Euclidean distances along multiple PC-s, therefore we introduced a novel approach, called PC50 clustering, to identify the most similar genomes. We clustered our genomes according to the pairwise weighed Euclidean distances of the first 50 PCA Eigenvalues and Eigenvectors (PC50 distances), where distances were calculated as follows: $\sqrt[3]{W1 * P1^2 + W2 * P2^2 + \dots + Wn * Pn^2}$, where W=Eigenvalue and P=Eigenvector.

As our PCA was confined to the HO dataset, this analysis was also done with the HO dataset. According to the Tracy-Widom statistics the first 110 PC vectors contained significant variation in our data set while the first 50 PCs accounted for 61.26% of the total variation. For Hierarchical Clustering we applied (ward.D2)⁷⁰ implemented in R 3.6.3.⁵⁹

Published genomes are commonly grouped according to their major genetic ancestry components, but members of the same group often vary in the proportion of these. In order to obtain genetically homogenous source populations for qpAdm, we also re-grouped relevant published samples with the same method. For this end we projected 2364 relevant published Eurasian ancient samples on the same modern PCA background described above and performed the same PC50 clustering.

As clustering splits down to the individual level, for obtaining groups it makes sense to cut the smaller branches. We cut the tree arbitrarily at a relatively great depth, considering the 50 deepest branches, where published “homogenous” groups were already divided into subclusters. If samples from published genetic groups fell into different clusters, we subdivided the original group according to the clusters. To create as homogenous groups as possible, in some cases we distinguished subgroups even within the same clusters, if they were separated by relatively large PC1 or PC2 distances. This did not bias the results, as theoretically subgrouping can be performed down to the individual level, if qpAdm could handle thousands of sources. This way we regrouped Late Bronze Age, Iron Age and Medieval samples published in Allentoft et al.,¹⁸ Damgaard et al.,²⁰ Gneccchi-Ruscione et al.,²¹ Jeong et al.,²² Wang et al.,²³ Järve et al.,²⁵ Unterländer et al.,⁷¹ and Krzewińska et al.⁷² (Data S3). We generally merged all samples within each group, but when many samples were available from a certain group, we left out the ones below 50 thousand SNPs corresponding to the HO dataset, and merged the remaining genomes with PASS quality assessment (described in the Allen Ancient DNA Resource datatable).

The hierarchical clusters based on the first PC50 distances are shown in Data S3. PC50 clustering identified the Avar_Asia_Core and Conq_Asia_Core populations as homogenous groups, and even Core1 and Core2 subgroups are separated to finer branches. Though representatives of the Eur_Core groups were identified in a preliminary distal qpAdm analysis, all five Eur_Core groups fall on separate clusters, with the exception of two samples (TMH-388 and TMH-199), which fall into the neighboring cluster. We found that the sensitivity of PC-50 clustering for grouping genomes according to similarity is comparable to the resolution of individual qpAdm.

Admixture modeling using qpAdm

We used qpAdm⁷³ from the ADMIXTOOLS software package⁶⁰ for modelling our ADMIXTUREs as admixtures of 2 or 3 source populations and estimating ancestry proportions. The qpAdm analysis was done with the HO dataset, as in many cases suitable Right or Left populations were only available in this dataset. We set the details:YES parameter to evaluate Z-scores for the goodness of fit of the model (estimated with a Block Jackknife).

As we tested a large number of source populations, testing every possible combination of sources (Left populations) and outgroups (Right populations) was impossible. Instead we ran the analysis just with source combinations of 2 and 3 (rank 2 and 3). As qpWave is integrated in qpAdm, the nested p values in the log files indicate the optimal rank of the model. This means that if p value for the nested model is above 0.05, the Rank-1 model should be considered.⁷³

To reveal past population history of the Test populations from different time periods, we run two separate qpAdm analysis. In the so-called “distal analysis” pre-Bronze Age and Bronze Age populations were included as sources. Next we run a so-called “proximodistal analysis”, in which just the most relevant distal sources were included in the Left population list, supplemented with a large number of post-Bronze Age populations. In latter runs potentially more relevant proximal sources competed with distant Bronze Age sources, and plausible models with distal sources indicated the lack of relevant proximal sources. In some cases, we used modern populations as sources, because the more relevant ancient sources were seemingly unavailable.

We performed a preliminary distal qpAdm analysis for each Eur-cline sample (data not shown), which supported PC50 clustering. As a result, we established five Eur-cline groups representing five different genome compositions. Next, we selected a few representatives from each group, with the most equivalent qpAdm models, and merged these genomes under the name of Eur_Core1 to Eur_Core5 respectively.

qpAdm analysis strategy

We must note that our Middle Age samples are highly admixed (Figure 2B; Data S5), likewise most of the Iron Age populations from which they possibly derived contain similar genomic components obtained from various admixture histories of related populations. In such cases qpAdm expectedly results in multiple feasible alternative models, most of which is very difficult to exclude. As we wished to include all relevant potential source populations in the analysis, despite their clearly similar genome histories, excluding suboptimal models was the largest challenge of the analysis. Thus, we set out to optimize our qpAdm strategy in multiple steps. Since our Test populations had largely different genome structures, we optimized the Right populations for each analysis.

In the first step we performed an iterative optimization of our Right populations, to exclude redundant, non discriminative Right populations, based on the log analysis of qpAdm for each run. Based on PCA, outgroup f3-statistics and unsupervised ADMIXTURE data, we began by assembling a large set of plausible pre-Bronze Age and early Bronze Age Right populations containing different ancestry components present in our Test populations. After several initial runs with a diverse set of Right populations, we collected the models with at least one significant Z-score from the detailed “gendstat” lines of the log files of all qpAdm models. We also counted how many models we would not reject if we excluded the f4-statistics with significant Z-scores of a given Right population. Based on this information we could test if all the Right populations were needed to reject the models. Then we repeated the qpAdm analysis with the optimally reduced Right populations until most Right populations were needed to reject models. As an important exception, we always kept Right populations that measured the main genetic components of our test population. Since all our Test populations were Eurasian samples we used a suitable outgroup, (Ethiopia_4500BP_published.SG) as Right Base throughout our analyses. As a result of Right optimization, a unique Right population set was used in each analysis, which are listed in details in Data S4, S7, S8, and S9, where qpAdm results are presented.

In order to further exclude suboptimal models, finally we applied the “model-competition” approach described in Narasimhan et al.³⁰ the following way: From the Left populations present in the plausible models we moved one at a time to the Right set and rerun qpAdm for each model. This was repeated with each Left population which appeared in any of the plausible models. As the best sources have the highest shared drift with the Test population, including these in the Right list is expected to exclude all models with similar suboptimal sources. This way we were able to filter out most suboptimal models and identify the most plausible ones, which were not excluded by any of the Right combinations. As each model-competition run gave a different p value with different standard deviations, we deemed it more informative to provide the maximum, minimum and average p values for the best final models, instead of the p values and standard deviations of the original models.

In some cases, model-competition excluded all 3 source models, indicating that additional sources are required for optimal modeling. However, running 4 source models with numerous sources is not feasible, thus we run 4 source models with a reduced set of Left populations. To select the best candidate subset of sources we evaluated the gendstat data from the log files of the 3 source models and identified populations which excluded the best models. Latter populations presumably had additional shared drift with the Test, not shared by any of the sources. Thus, in the rank 4 models we included just the best sources from the rank 3 models, plus their excluding populations.

Many of our samples were part of genetic clines between East and West Eurasia. In order to reveal the genetic ancestry of individual samples within genetic clines the identified genetic groups at the eastern and western extremes were also added to the Left-populations as sources. Many of the samples within clines could be modelled as simple two-way admixture of these two populations, or three-way admixtures with a third source. The remaining individuals were considered genetic outliers, which were modelled from different sources. As three source modeling of large number of samples from large number of sources is not feasible, we decreased computation time by decreasing the sources (removing the most redundant ones), and running constrained 3 source models by fixing one source at a time, and swapping only the rest of the Left populations. We repeated these constrained runs in 3 combinations, once fixing Eur_Core, second fixing Conq_Asia_Core and third fixing Avar_Asia_Core. Finally, we compared the 3 different models for each sample and selected the ones with best p values. Obtaining the same high p value models from two different constrained runs, one with Eur_Core fixed, the other with Asia_Core fixed, rendered it very likely that the optimal model was found. In cases when Eur_Core fixed, and Asia_Core fixed runs gave different models, but all Asia_Core fixed models indicated the presence of

Eur_Core, we accepted the best Eur_Core fixed models. Finally, we also tested whether samples at the Asian side of the cline require entirely different sources by running unconstrained 3 source models for these few samples.

We noticed that model-competition cannot exclude alternative qpAdm models with similar minor components, therefore a given source is best identified from samples which carry it in large fraction. We had several within cemetery PCA clines with varying fractions of the same components, as nearly each member could be modelled from individuals at the extremes of the cemetery cline (Data S8E). In these cases the best Asian source could be accurately identified from the unconstrained 3 source models, which also applies to all members of the cemetery cline.

We present qpAdm results in the following format in Data S4, S7, S8, and S9: All passing models are shown which received significant p values after model competition runs, or the best models if none of the models passed (unless indicated otherwise). Column LEFT lists the source populations which were tested in each 2-way or 3-way combinations for the given Test populations. Column RIGHT lists the reference populations used in the given qpAdm. The "total" column indicates the number of competition runs repeated for each plausible model. Column "valid" shows how many times the given model passed, while "excluded" shows how many times it was excluded. In latter case the excluding extra Right population is shown in the "excluding pop" column. "neg" column indicates the number of models in which any of the sources produced a negative fraction. "nested P" column indicates the number of models with significant (>0.05) nested p values, when the Rank-1 model is more plausible. "max p" indicates the maximum p value obtained from the competition runs, while "max P right pop" shows the extra Right population in this model. "min P" and "min P Right pop" means the same for the lowest p value model, while "avg P" means the average p value of the valid models. Models are arranged according to max p and min p values.

f3-statistics

Outgroup f3-statistics is suitable to measure shared drift between two test populations after their divergence from an outgroup⁷⁴ thus providing a similarity measure between populations. We measured the shared drift between the identified homogeneous new genetic groups in our sample set and all published modern and ancient populations, to identify populations with shared evolutionary past. As an outgroup we used African Mbuti genomes and applied ADMIXTOOLS⁶⁰ to calculate f3 statistics. This analysis was also done with the HO dataset, and we removed populations below 40 thousand overlapping markers.

Admixture f3-statistics in the form $f3(\text{Test}; X, Y)$ can be used to identify potential admixture sources of the Test population,⁶⁰ and most negative f3 values indicate the major admixture sources. We used the qp3Pop program of ADMIXTOOLS with the *inbreed: YES* parameter. All population combinations are listed in Data S6C and S6D.

Two dimensional f4-statistics

To measure different levels of bi-directional gene flow into populations with shared genomic history Narasimhan et al.³⁰ applied a so called "Pre-Copper Age affinity f4-statistics", with a 2-dimensional representation of the f4 values from two related statistics. This way populations or individuals with significantly different proportions of ancestry related to the two sources can be visualized. We applied this 2-dimensional f4-statistics to measure gene flow from multiple sources into members of the same population. This way we could explore the fine substructure of populations and identify potential sources of the gene flow. To calculate f4-statistics we used the qpF4ratio from ADMIXTOOLS.⁶⁰ We run two-dimensional f4-statistics with the 1240K marker set.

Dating admixture time with DATES

The DATES algorithm³⁰ was developed to infer the date of admixture, and this software was optimized to work with ancient DNA and single genomes. As qpAdm often revealed that a two- or three-way admixture well explained the genome history of the studied population, we used DATES to determine admixture time. As we used modern populations with DATES, this analysis was also confined to the HO dataset.