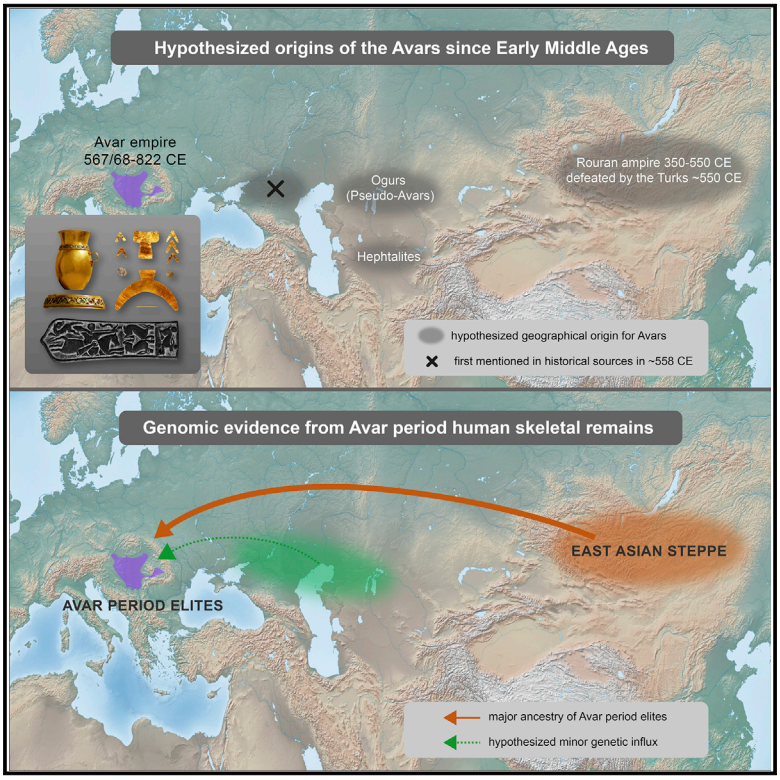
Cell

Ancient genomes reveal origin and rapid *trans*-Eurasian migration of 7th century Avar elites

Graphical abstract



Highlights

- Long-distance and rapid *trans*-Eurasian migration during the 7th century Avar period
- Striking genetic similarity between early Avar elites and the Rouran in Mongolia
- Substantial genetic variation mirroring social and micro-geographic structure
- High eastern Eurasian ancestry maintained in the Avar period elites for 200 years

Authors

Guido Alberto Gnecchi-Ruscone, Anna Szécsényi-Nagy, István Koncz, ..., Zuzana Hofmanová, Choongwon Jeong, Johannes Krause

Correspondence

guido_gnecchi@eva.mpg.de (G.A.G.-R.), cwjeong@snu.ac.kr (C.J.), krause@eva.mpg.de (J.K.)

In brief

The Avars were a mysterious population that settled the Carpathian Basin in 567/68 CE, and their origins have remained enigmatic. Genomic analyses of 66 pre-Avar and Avar-period individuals, integrated with archaeological and historical data, suggest that Avar elites underwent a long-distance, *trans*-Eurasian migration from the East Asian steppe.

Gnecchi-Ruscone et al., 2022, Cell *185*, 1402–1413
April 14, 2022
https://doi.org/10.1016/j.cell.2022.03.007

CellPress

CellPress
OPEN ACCESS

Cell

Article

Ancient genomes reveal origin and rapid *trans*-Eurasian migration of 7th century Avar elites

Guido Alberto Gnecchi-Ruscone,[1,32,*] Anna Szécsényi-Nagy,[2,32] István Koncz,[3] Gergely Csiky,[4] Zsófia Rácz,[3] A.B. Rohrlach,[1,5] Guido Brandt,[6] Nadin Rohland,[7,8] Veronika Csáky,[2] Olivia Cheronet,[9] Bea Szeifert,[2] Tibor Ákos Rácz,[10] András Benedek,[11] Zsolt Bernert,[12] Norbert Berta,[13] Szabolcs Czifra,[12] János Dani,[14] Zoltán Farkas,[13] Tamara Hága,[14] Tamás Hajdu,[15] Mónika Jászberényi,[10] Viktória Kisjuhász,[16] Barbara Kolozsi,[14] Péter Major,[13] Antónia Marcsik,[17]

(Author list continued on next page)

[1]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[2]Institute of Archaeogenomics, Research Centre for the Humanities, Eötvös Loránd Research Network, 1097 Budapest, Hungary
[3]Institute of Archaeological Sciences, ELTE Eötvös Loránd University, 1088 Budapest, Hungary
[4]Institute of Archaeology, Research Centre for the Humanities, Eötvös Loránd Research Network, 1097 Budapest, Hungary
[5]ARC Centre of Excellence for Mathematical and Statistical Frontiers, School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia
[6]Max Planck Institute for the Science of Human History, 07745 Jena, Germany
[7]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[8]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[9]Department of Evolutionary Anthropology, University of Vienna, 1030 Vienna, Austria
[10]Ferenczy Museum Center, 2000 Szentendre, Hungary
[11]Móra Ferenc Museum, 6720 Szeged, Hungary
[12]Hungarian National Museum, 1113 Budapest, Hungary
[13]Salisbury Ltd., 1016 Budapest, Hungary
[14]Déri Museum, 4026 Debrecen, Hungary
[15]Dept. of Biological Anthropology, Eötvös Loránd University (ELTE), 1117 Budapest, Hungary
[16]Aquincum Museum and Archaeological Park, 1031 Budapest, Hungary
[17]Dept. of Biological Anthropology, Szeged University, 6701 Szeged, Hungary
[18]Katona József Museum, 6000 Kecskemét, Hungary
[19]Department of Art History, Istanbul Medeniyet University, 34720 Istanbul, Turkey
[20]Research Centre for the Humanities, Eötvös Loránd Research Network, 1097 Budapest, Hungary
[21]Wosinsky Mór Museum, 7100 Szekszárd, Hungary

(Affiliations continued on next page)

SUMMARY

The Avars settled the Carpathian Basin in 567/68 CE, establishing an empire lasting over 200 years. Who they were and where they came from is highly debated. Contemporaries have disagreed about whether they were, as they claimed, the direct successors of the Mongolian Steppe Rouran empire that was destroyed by the Turks in ~550 CE. Here, we analyze new genome-wide data from 66 pre-Avar and Avar-period Carpathian Basin individuals, including the 8 richest Avar-period burials and further elite sites from Avar's empire core region. Our results provide support for a rapid long-distance *trans*-Eurasian migration of Avar-period elites. These individuals carried Northeast Asian ancestry matching the profile of preceding Mongolian Steppe populations, particularly a genome available from the Rouran period. Some of the later elite individuals carried an additional non-local ancestry component broadly matching the steppe, which could point to a later migration or reflect greater genetic diversity within the initial migrant population.

INTRODUCTION

Long-distance migration to Europe is rarely reported in historical sources. The Avars, who arrived in the Carpathian Basin from the Central Eurasian steppes in 567–568 CE, are an iconic exception. Their empire, ruled by a khagan, dominated eastern Central Europe for over 200 years, until it was overcome by the Franks around 800 CE (Curta, 2021; Daim, 1992; Pohl, 2018; Szádeczky-Kardoss,

CellPress
OPEN ACCESS

Bernadett Ny. Kovacsóczy,[18] Csilla Balogh,[19] Gabriella M. Lezsák,[20] János Gábor Ódor,[21] Márta Szelekovszky,[14] Tamás Szeniczey,[15] Judit Tárnoki,[22] Zoltán Tóth,[23] Eszter K. Tutkovics,[24] Balázs G. Mende,[2] Patrick Geary,[25] Walter Pohl,[26,27] Tivadar Vida,[3] Ron Pinhasi,[9] David Reich,[7,8,28,29] Zuzana Hofmanová,[1,30] Choongwon Jeong,[31,*] and Johannes Krause[1,33,*]

[22]Damjanich Museum, 5000 Szolnok, Hungary
[23]Dobó István Museum, 3300 Eger, Hungary
[24]Rétközi Múzeum, 4600 Kisvárda, Hungary
[25]Institute for Advanced Study, Princeton, NJ 08540, USA
[26]Institute for Medieval Research, Austrian Academy of Sciences, 1020 Vienna, Austria
[27]Institute of Austrian Historical Research, University of Vienna, 1010 Vienna, Austria
[28]Department of Human Evolutionary Biology, Cambridge, MA 02138, USA
[29]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
[30]Department of Archaeology and Museology, Faculty of Arts, Masaryk University, 60200 Brno, Czech Republic
[31]School of Biological Sciences, Seoul National University, 08826 Seoul, Republic of Korea
[32]These authors contributed equally
[33]Lead contact
*Correspondence: guido_gnecchi@eva.mpg.de (G.A.G.-R.), cwjeong@snu.ac.kr (C.J.), krause@eva.mpg.de (J.K.)
https://doi.org/10.1016/j.cell.2022.03.007

1990). The Byzantine texts agree that their move had been triggered by the rise of the first Turkic khaganate in the 550s, centered in what is now Mongolia, when Turks destroyed an empire called Rouran by its Chinese neighbors (Kradin, 2005). However, the texts do not agree on who these Avars were, or where exactly they came from. In fact, the Turks claimed that they were only Pseudo-Avars who had appropriated the prestigious name Avars and the lofty title of khagan but were in reality Ogurs, a Turkic-speaking people in western Central Eurasia. While we can conclude that the Rouran most likely called themselves Avars, to what extent the European Avars were descended from them has been debated (Dobrovits, 2003; Pohl, 2018). Here, we present genomic data that provide a new basis to reconstruct the early medieval long-distance movements of steppe peoples (Alves et al., 2016; Damgaard et al., 2018; Gnecchi-Ruscone et al., 2021) and an opportunity to integrate genetic, historical, and archaeological evidence (Savelyev and Jeong, 2020).

Before the Avars arrived, the Romans had occupied the western part of the Carpathian Basin and the Sarmatians the eastern part (c. 1–400 CE). The Romans were replaced by the short-lived empire of the Huns (400–455 CE), and by diverse Germanic-speaking groups: Goths and Longobards in Pannonia, Gepids along the Tisza (400 to c. 568). In 567/68, the Longobards destroyed the Gepid kingdom and moved to Italy, while the Avars conquered the Carpathian Basin and its local population (Pohl, 2018). This study focuses on this momentous change and its genetic impact.

Previous studies utilizing uniparental markers (Csáky et al., 2020; Neparáczki et al., 2019) have provided suggestive genetic evidence, but genome-wide data for reconstruction of the origins of the Avar-period population are missing. We use nuclear DNA to gain insights into the following questions: (1) can the origin of a core Avar group from eastern Central Asia be confirmed from their genomic profile? (2) Were the elites of the newly arrived steppe warriors genetically homogeneous or did they have mixed ancestries? (3) How do the elites relate to the preceding local population?

The rich archaeological material of the Avar period in the Carpathian Basin (late 6th c.–early 9th c. CE) consists of c. 600 settlements and c. 100,000 excavated burials, which show great heterogeneity, especially in the early phases (Bálint, 2019; Daim, 2003; Vida, 2016). We generated genome-wide data for 66 ancient individuals, covering both the pre-Avar (17 in a Sarmatian and one in a Hunnic context, 4th–5th c. CE) and Avar periods (N = 48, 7th–8th century CE) (Figure 1). The latter originate from three distinct regions. First, 25 individuals were collected from the Danube-Tisza Interfluve (DTI). Of these, 8 were high-status male burials (Bócsa-Kunbábony group) dated to the second third of 7th c. CE, after the failed siege of Constantinople in 626 CE (Hurbanič, 2019), containing exquisite gold- and silver-decorated weapons and belts, various insignia, and other prestige objects with Inner Asian parallels (Bálint, 2019; Vida, 2016) The remaining 17 individuals were taken from contexts with direct contact to this elite group (Data S1).

We also considered a group of burials mostly found east of the Tisza river (Transtisza), an assumed secondary power center, which share numerous cultural similarities with the 6th–7th CE population of the eastern European steppes. These are characterized by solitary burials or small burial grounds, the deposition of partial or complete animals, and ledge or end niche graves. Twelve of the 17 samples from the region belong to this "Transtisza group" (Gulyás, 2016; Lőrinczy, 2017). The remaining 6 Avar-period individuals were collected from neighboring regions in the Carpathian Basin and reflect the heterogeneity of archaeological material and the burial practices during the Avar period, including two richly furnished elite burials from Transdanubia (Kölked), where the archaeological record indicates a variety of local groups showing strong connections to the Mediterranean as well as the Merovingian world in western Europe (Vida, 2018).

RESULTS

Ancient DNA dataset and quality control

After applying an in-solution enrichment protocol for 1.24 million informative single-nucleotide polymorphisms (SNPs) (Fu et al., 2013a; Mathieson et al., 2015), we sequenced the enriched libraries to a median 2.9× coverage for the "1,240K" target sites, covering 13,749–1,119,583 target SNPs at least once (median

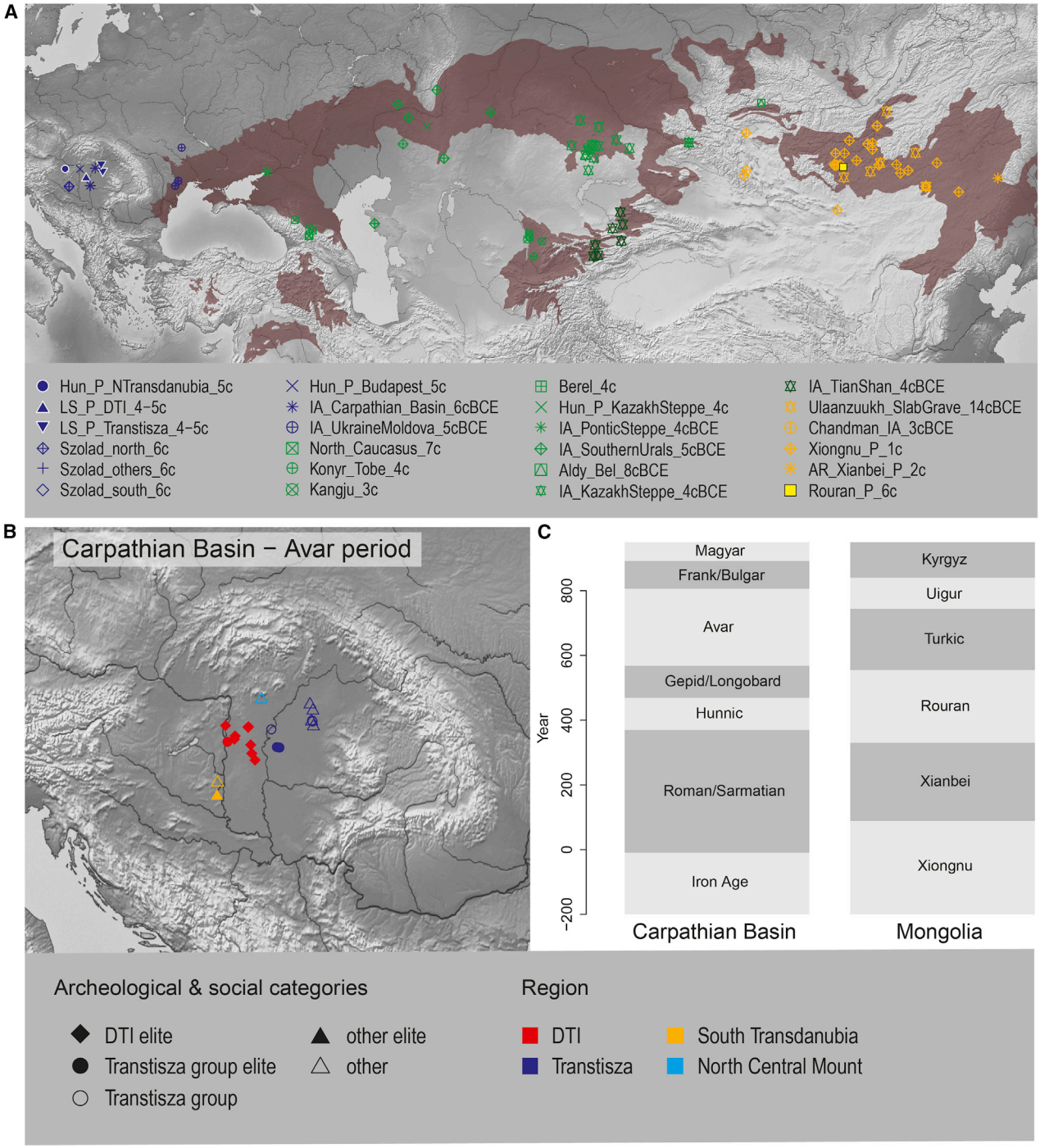


Figure 1. Geographic and temporal locations of ancient individuals in this study

(A) A map of Eurasia with geographic coordinates of the ancient individuals analyzed in this study marked by color-filled shapes. Dark yellow shades mark the steppe ecoregion. Newly produced genomes from the Carpathian Basin pre-Avar period are highlighted with white outlined symbols.

(B) A zoom-in map of the Carpathian Basin showing geographic coordinates of the newly analyzed ancient samples from the Avar period with symbols referring to specific archaeological and social categories and colored according to the regions, as defined in the bottom left and right legend respectively.

(C) A rough timeline of Carpathian Basin and Mongolia from 200 BC to 950 AD.

See also Table S1.

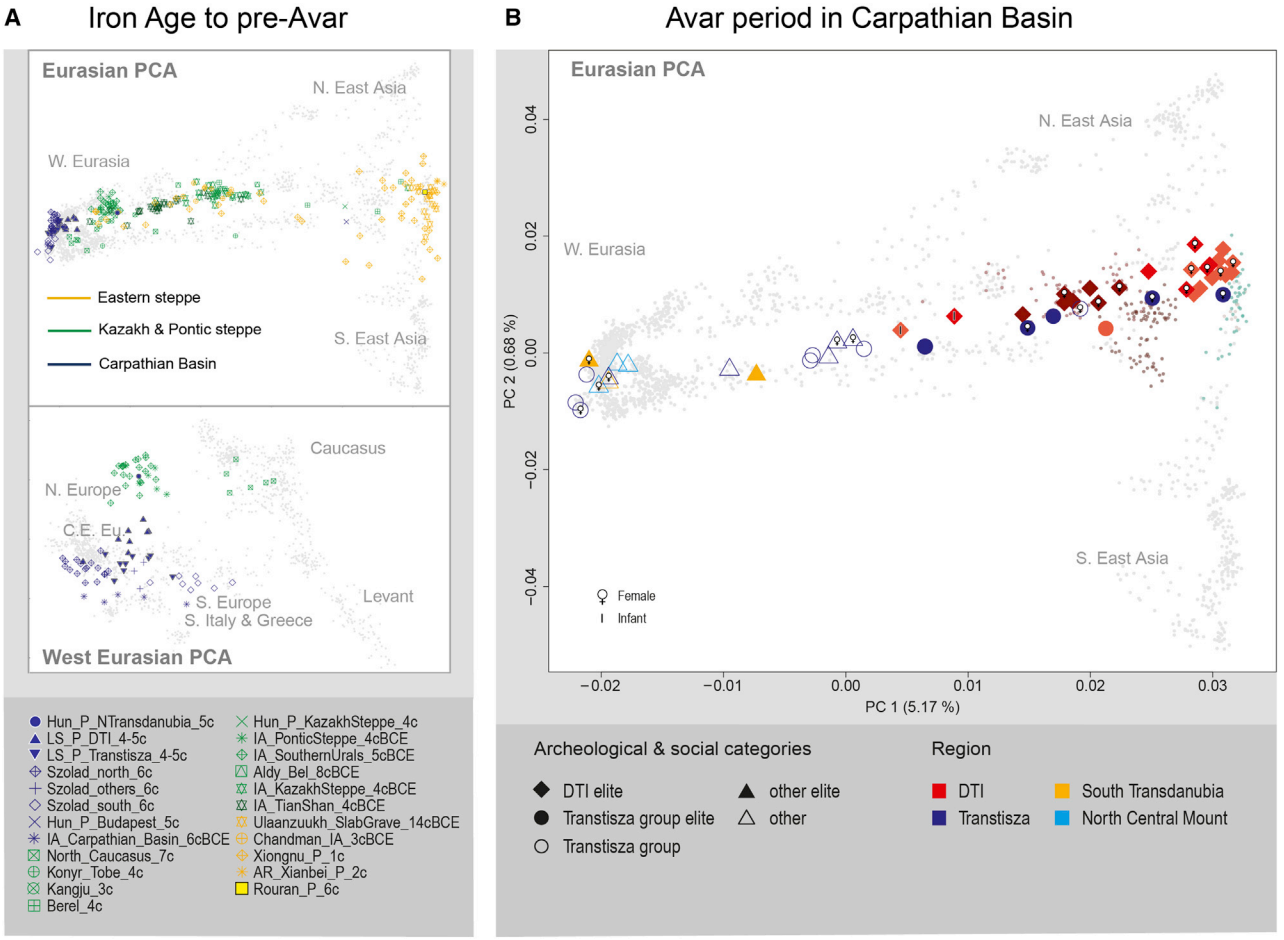


Figure 2. Principal component analyses

(A) Pre-Avar Eurasian PCA (top) and west Eurasian PCA (bottom). New data are highlighted with white outlined and filled symbols.

(B) Eurasian PCA of newly produced Avar period individuals. Colors refers to key regions within the Carpathian Basin. Filled symbols are individuals retrieved from elite contexts. Specific archaeological categories discussed in the text are shown with different symbol shapes.

See also Figure S1A.

887,064 SNPs). We find low mitochondrial contamination estimates (1%) for 61 of 66 individuals (one 3% and four unestimated) and low nuclear contamination estimates (<1%) for 36 of 38 males (Table S1). The remaining individuals do not have sufficient mitochondrial or X chromosome coverage to properly estimate contamination. We carry all 66 individuals into the downstream analysis considering that individuals with no contamination estimate still show similar genetic profiles as others with low contamination (including also a first-degree relative pair) and that all individuals have clear ancient DNA-damage patterns and expected sex chromosomes coverage ratios (STAR Methods; Table S1). To address specific questions detailed below we imputed the missing position of the 1,240K panel in order to obtain diploid genotype calls that we could statistically phase and perform haplotype-based local ancestry analyses following a thorough evaluation of the success of the imputation and exclusion of poorly imputed samples (i.e., samples with average genome-wide coverage <1.43×; STAR Methods; Figure S2A). To identify if more than just mean coverage was an important parameter when considering which samples we could include for imputation, we modeled the change in pairwise mismatch rate (pMMR) using mean coverage and contamination as predictor variables (see STAR Methods). We found that while mean coverage was a significant predictor, contamination was not. Hence, spurious signals of ancestry induced by minor levels of contamination did not significantly affect measures of pMMR before and after imputation, and we expect that the very minor levels of damage found between the two terminal bases of each read will also have no significant effect.

The genomic structure of the pre-Avar-period population

Principal component analysis (PCA) reveals markedly different genetic profiles between pre-Avar- and Avar-period individuals from the Carpathian Basin (Figures 2A and 2B, respectively). With the exception of the only two Hunnic-period genomes available (Hun_P_Budapest_5c, Hun_P_NTransdanubia_5c),

one newly generated and one previously published (Gnecchi-Ruscone et al., 2021), all pre-Avar individuals from the Carpathian Basin fall in the genetic variability of west Eurasians. They mostly overlap present-day central and eastern Europeans, although a few individuals align with southern Europeans, especially a previously described Longobard-period group (Szolad_south_6c) that overlaps present-day southern Italians and Greeks (Figures 2A and S1; Amorim et al., 2018). Comparing within the Carpathian Basins' ancient populations available (hereafter refer to as "local" populations), the 16 newly analyzed 4th–5th century CE late Sarmatian/Hunnic-period individuals fall close to but deviate from the 8th–4th century BCE individuals of the region (Carpathian_Basin_IA) and are closer but only partially overlapping to the 6th century CE Longobard-period ones (Amorim et al., 2018). Among the late Sarmatian-period individuals the ones from Transtisza (LS_P_Transtisza_4-5c) overlap with the Szolad_others_6c group, the ones from DTI (LS_P_DTI_4-5c) are still closest to LS_P_Transtisza_4-5c individuals but shifted toward Hun_P_NTransdanubia_5c and the Iron Age groups from the steppes (IA_PonticSteppe_4cBCE and IA_SouthernUrals_5cBCE) (Damgaard et al., 2018; Gnecchi-Ruscone et al., 2021; Krzewińska et al., 2018; Unterländer et al., 2017). These observations on PCA are supported and confirmed by the *qpWave*/*qpAdm* modeling of LS_P_Transtisza_4-5c and LS_P_DTI_4-5c (Figure S5C).

The genomic structure of the Avar-period population

Contrary to the preceding periods, the Avar-period individuals show considerable genetic variability as the sampled individuals are spread along the entire cline from West Eurasian to Northeast Asian populations (Figure 2B). Despite this overall heterogeneity, there are clear patterns of genetic substructure corresponding both with Carpathian Basin's geography and social-archaeological categories (Figure 2B). All of the individuals from the DTI elite sites have Northeast Asian ancestry profiles falling along a genetic cline of present-day populations from the Altai to Mongolia and to the Amur River Basin (Figure 2B). They also broadly fall within the variability of ancient individuals associated with some of the main late Iron Age to early Medieval eastern steppe archaeological cultural horizons (Figures 2A and 2B). Among them, a wider group consists of 3rd BCE–1st c. CE Xiongnu period individuals from the Mongolian plateau (N = 46, Xiongnu_1c) that are overall highly heterogeneous and have been previously grouped into three clusters, based on their genetic profiles (Jeong et al., 2020). Other individuals include the 1st–3rd CE Xianbei period individuals from the Amur River Basin (N = 3, AR_Xianbei_P_2c) (Ning et al., 2020) and the Altai (N = 6, Berel_4c) (Gnecchi-Ruscone et al., 2021); two Hun-period individuals, one individual from the early 5th century Carpathian Basin, one individual from the 4th century Kazakh steppe (Hun_P_Budapest_5c and Hun_P_KazakhSteppe_4c) (Gnecchi-Ruscone et al., 2021; Nagy, 2010) and one 6th c. CE Rouran-period individual from present-day Mongolia (Rouran_P_6c) (Li et al., 2018). All of these individuals, albeit variably mixed with other sources, have been shown to trace their eastern Eurasian ancestry component to a genetic profile referred to as the "ancient northeast Asians" (ANA) (Jeong et al., 2019).

The genetic analysis does not use cultural assignments or chronological information, but nonetheless clusters Avar-period DTI elite individuals according to the Avar chronology (early, middle, and late). All of the early-Avar-period individuals (DTI_early_elite), except for an infant and a burial with typical characteristics of the Transtisza group (Figure 2B), form a tight cluster with a high level of ANA ancestry. They are located between present-day Mongolic- (e.g., Buryats and Khamnigans) and Tungusic/Nivkh-speaking populations (e.g., Negidals, Nanai, Ulchi, and Nivkhs) together with the only available ancient genome from the Rouran-period Mongolia and are close to the three AR_Xianbei_P_2c individuals in the PCA (Figures 2A and 2B). Three out of six middle Avar-period individuals (DTI_middle_elite) fall within the DTI_early_elite cluster, two (DTI_middle_elite_o) slightly depart from it and the last one (a child burial) cluster in an intermediate position along the Eurasian PC1. The late Avar-period individuals (DTI_late_elite) are distinct from the earlier ones as they are all shifted toward west Eurasia in PCA (Figure 2B).

Modeling the eastern steppe ancestry of the elites in the core of the Avar empire

Genetic ancestry modeling performed with the *qpWave*/*qpAdm* framework confirmed that the DTI_early_elite individuals and the three DTI_middle_elite can be modeled as carrying 88%–98% of their ancestry from an ANA-related gene pool (~90% on average) while the DTI_late_elite individuals carry between 70% and 80% of such an ancestry source (Figures 3 and S1).

We chose AR_Xianbei_P_2c from the Mogushan archaeological site (n = 3), contextually dated to 50–250 CE, as a proxy for ANA due to its relatively high coverage, it being closer in time and closer genetically to the Avar-period individuals with the easternmost ancestry pattern and because of an historical connection between the Xianbei and Rouran Empires (Golden, 2013). Furthermore, local ancestry analyses performed on the imputed-phased data revealed that when masking the genomic regions of western ancestry, all of the Avar-period elite individuals cluster tightly together close to the position of AR_Xianbei_P_2c individuals (Figure S2C). Nevertheless, replacing AR_Xianbei_P_2c with other ancient eastern Eurasian populations that carry high proportions of ANA ancestry also yields fitting models with no qualitative difference (Table S2). Among them, Rouran_P_6c yields fitting models as a single source of ancestry for many of the DTI early/middle elites, although these models have low statistical power due to low coverage of the Rouran_P_6c individual (Table S2). In turn, Rouran_P_6c can be modeled with the same two-way sources at comparable proportions as the DTI early elite individuals (Figure 3).

The remaining ancestry, ~10% on average for the DTI early and middle elite and 20%–30% in DTI late elite individuals, comes from a source carrying higher western Eurasian ancestry (Figures 3 and S1; Table S2). With the exception of two DTI late individuals, a broad range of ancient populations equally provide working models, both from the Pontic steppes and 4th–6th century Carpathian Basin groups (Table S2). We replicated the *qpWave*/*qpAdm* modeling grouping the individuals based on chronology to enhance statistical power (with the exclusion of

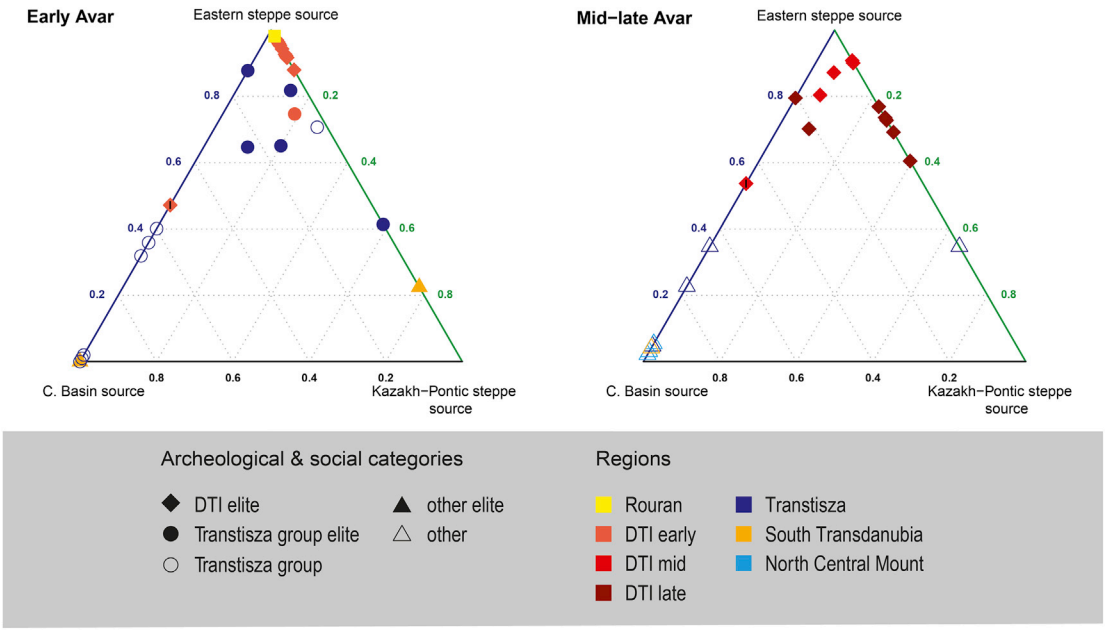


Figure 3. Ancestry deconvolution performed with *qpWave*/*qpAdm*

Representative two- or three-way admixture models for the Avar period individuals. Early Avar-period on the left and middle-late Avar period on the right. The figure reports the overall best models resulting from the evaluation of all the individual-based, group-based and local ancestry-based qpAdm models following the rationale detailed in the STAR Methods section. Three possible sources of ancestry are tested, an eastern Asian steppe source (AR_Xianbei_P_2c is used for all models reported in the figure), a Carpathian Basin source (blue side, represented by either one among the three Longobard-period Szólád groups or the Sarmatian-period group), and a Pontic-to-Kazakh steppe source (green side, represented by either one of the Iron Age groups from the steppe or the North_Caucasus_7c). In Figures S1, S4, and S5 are shown the specific sources within these geographic categories that provided fitting models for each of the tested individuals.

See also Figures S1, S4, and S5 and Tables S2 and S4.

the outliers and one from each pair of related individuals). DTI_early_elite provides a good fit only with the Iron Age IA_Chandman_3cBCE from western Mongolia and a marginal non-fit with IA_SouthernUrals_5cBCE (χ^2 p = 0.07 and 0.049, respectively; Table S3). DTI_late_elite provides working models only with western proxies closer in time (1st millennium CE), but the geographic source could not be fully resolved since local groups as well as a north Caucasus group from the 7th century (North_Caucasus_7c) and a southern Kazakhstan group from the 4th century (Konyr_Tobe_4c) provide fitting models (Table S3). To further investigate the issue of local versus steppe/Caucasus origin of the western source in DTI_late_elite, we applied three-way competing models (STAR Methods), contrasting pairwise combinations of local versus non-local sources (Figure S1D). These revealed that for most individuals (DTI_late_elite1) except two (DTI_late_elite2) a non-local source was preferred over a local one, for which North_Caucasus_7c or present-day North Ossetians provided the best fits. Group-based two-way models on these two subgroups confirmed this pattern (Tables 1 and S3), thus revealing higher resolution to distinguish between western sources with a two-way model. We also repeated *qpAdm*/*qpWave* ancestry modeling on the imputed-phased data after performing local ancestry analysis and masking the eastern Eurasian genomic regions (i.e., analyzing only the western ancestry of the tested individuals; STAR Methods and Figures S2B and S2C). Both individual and group-based analyses were consistent with the findings obtained on the pseudohaploid data (Figure S2B). A caveat regarding the non-local source identified, however, is that no data from the core area of 4th–6th c. CE Hunnic realm north of Persia are yet available to test for more-specific alternative geographic sources.

The inferred dates of admixture in the DTI elites corroborate the distinction between the early and late elite individuals: the early period elites, as well as the three middle period elite individuals, present older admixture dates compatible with the date estimated for the Rouran-period individual, while the late elites and the two middle elite outliers show substantially more recent admixture dates (Figures 4 and S3). Per-individual admixture date estimates for the early/middle elites range from the 4th century BCE to the 3rd century CE and the group-based one falls in between, around the 1st century BCE (Figures 4 and S3). By contrast, the late elite individuals show more recent admixture date estimates that fall within the Rouran or early European Avar periods, and more precisely toward the end of the Rouran period in the group-based analysis (Figures 4 and S3).

Finally, even if we found a considerable number of closely related individuals among the DTI elites within the same sites and within both early and late periods (five 1st and one 2nd degree pairs, including a trio; Data S1), they show no signs of recent inbreeding (except for one individual) (Figure S4). Nevertheless, the East Central Asian paternal lineage N-F4205 that was

Table 1. p values of the group-based *qpWave/qpAdm* models of the two DTI_late_elite groups

	Second sources (the first is DTI_early_elite) of two-way *qpAdm* models with OG1			
Targets	North_Caucasus_7c	Szolad_north_6c	LS_P_Transtisza_4–5c	LS_P_DTI_4–5c
DTI_late_elite1	p value = 0.66[a]	p value = 5.5×10^{-8}	p value = 0.0004	p value = 0.0007
DTI_late_elite2	p value = 0.008	p value = 0.701[a]	p value = 0.89[a]	p value = 0.57[a]

DTI_late_elite1 (A1809, A1813, A1814, A1815, I18222) include individuals with predominantly non-local signal while DTI_late_elite2 (A1810, I18225) include individuals with predominantly local signal from the three-way models in Figure S1D and Table S3.
[a]p values > 0.05

detected previously as typical for the elite males buried in this region (Csáky et al., 2020) shows continuity throughout the middle and late Avar periods (Table S1). All twenty Avar-period males from the DTI carried the N1a1a1a1a (N-F4218) lineage, and all but one could be assigned to the N-F4205 sub-branch, typical for present-day Mongolian and Transbaikalian populations (Ilumäe et al., 2016). The mtDNA diversity has instead been shown to be remarkably higher (Csáky et al., 2020). This, coupled with the absence of genomic signs of inbreeding or of evidently reduced population sizes (Figure S4), could suggest patrilocality/patrilineality practices as well as higher female exogamy that would have prevented inbreeding.

The heterogeneous ancestry in the regions surrounding the Avar empire's core

The remaining 23 individuals come from the neighboring regions surrounding the DTI. Compared with the elites of the Avar empire's core region, they are widely spread in PCA space along a west-to-east ancestry cline, from the local preceding gene pools (represented by Sarmatian- and Longobard-period individuals) to the genetically easternmost DTI_early_elite cluster (Figures 2B and S1). Nevertheless, only individuals retrieved from Transtisza group burials carry >50% of ANA-related ancestry (7 out of 13 Transtisza group individuals; Figures 3 and S5A; Table S4). All except one were sampled from the two elite sites (Figure S5A). The remaining Transtisza group individuals show little to no detectable ANA-related admixture (Figures 3 and S5A). However, two Transtisza group individuals with no ANA ancestry show a higher affinity toward Mediterranean populations clustering with present-day Sicilians and Maltese, and plot at the very end of the Szolad_south_6c genetic cluster (Figure S1). For this reason, they cannot be fully modeled with any of the sources used in our study, although the Szolad_south_6c provides the models with best fit as a single source when the different local preceding groups are contrasted in competitive models (Figure S5B). The genetic heterogeneity of Transtisza group individuals is remarkable because this population has mostly been ascribed to immigrants from the Pontic steppes (Gulyás, 2016; Lőrinczy, 2017). It is likely due to varying degrees of admixture at different time points in the past (Figures 4 and S3): the two individuals with the highest ANA ancestry, indistinguishable from the DTI early elites, show similarly old admixture dates, while the rest with more western Eurasian ancestry have younger and more variable admixture dates ranging from the 1st century CE to just a few generations before (Figures 4 and S3). This suggests that despite a common cultural practice, individuals with very different genetic backgrounds were inhumated in these types of cemeteries. The last two elite-related individuals included in our dataset are the ones found at the early Avar-period site of Kölked in south Transdanubia. They show two very different genetic profiles as the female (A1825) falls on top of the preceding local population, while the male (A1824) has a unique genetic ancestry with respect to all of the other individuals described so far, showing both a shift toward east Eurasia as well as toward the Iranian/Caucasus-related gene pool, plotting on top of present-day Tajiks and close to ancient Caucasus steppes such as North_Caucasus_7c, Konyr_Tobe_4c and Kangju_3c (Figures 2B and S1). In agreement with this observation, A1825 can be modeled as deriving 100% of its ancestry from local preceding sources, while A1824 can only be modeled as a mixture between a ~20% ANA source (e.g., AR_Xianbei_P_2c) and ~80% North_Caucasus_7c, or present-day North Ossetians (Figures 3 and S5A; Table S4). The 5th CE admixture date obtained for this individual corroborates the interpretation of a non-local-recently admixed ancestry (Figures 4 and S3). The remaining 9 late-Avar-period individuals show minor (<40%) to almost negligible (<5%) admixture with ANA-related sources, while the major ancestry component can be approximated by one of the preceding local groups for most of the individuals (Figures 3 and S5A; Table S4). Admixture dates confirm a general pattern of more recent admixture occurring mostly during the earlier phases of the Avar empire although some slightly earlier dates, in the 5th century CE, could point to admixture events with eastern sources that occurred during the Hunnic period (Figures 4 and S3). Furthermore, our data suggest mostly unidirectional gene flow as, with some exceptions previously described (e.g., the two child/infant burials; Figure S1C), most of the recently admixed individuals are found in non-elite sites.

DISCUSSION

Our results provide robust genetic support for the Northeast Asian ancestry of the Avar-period elite in the core region of the Avar empire (DTI) from the middle third of the 7th CE to the early 8th CE Carpathian Basin (early to middle Avar period). We show a striking genetic match with a Rouran-period individual as well as with ancient individuals from Xiongnu and especially Xianbei periods from the eastern Asian steppe. During the late Avar period, we observe a shift among the elite in the Avar core area toward a more recently admixed ancestry. Even if late Avar individuals still preserve a predominant northern East Asian component, the western Eurasian source that best fits the remaining 20%–30% of their ancestry is mostly a non-local one (i.e., it does not match the gene pools of the available preceding Carpathian Basin populations). Instead, it rather

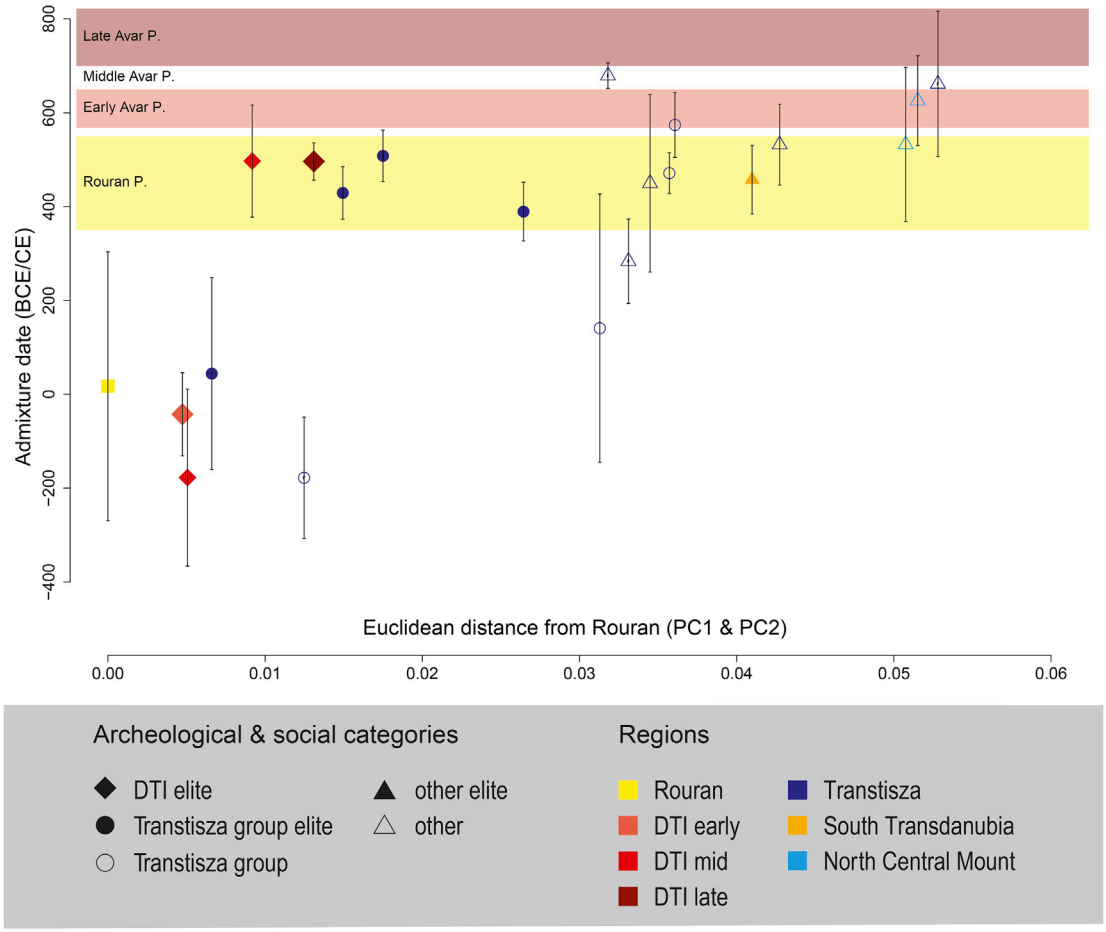


Figure 4. Admixture dating obtained with DATES

Admixture dates and respective error bars obtained with DATES (between a western and eastern Eurasian source) plotted against the Euclidean distance from PC1 and PC2 coordinates of the Rouran_P_6c genome (PCA of Figure 2B). The colored bands mark the chronology of the three Avar and Rouran periods, used to derive the dates of admixture from the estimated number of generations since admixture. All dates shown are individual-based except for the DTI elites (early, middle, and late) for which the group-based date estimates are shown in the plot and therefore the individuals' median Euclidean distance from Rouran_P_6c are used in this case. See also Figure S3.

matches the steppes north of the Caucasus, although the scarceness of comparative data from the steppe in the 1st millennium CE calls for a future investigation of possible better alternative sources.

This non-local admixture together with the retention of a high level of eastern ancestry in the late elites and the absence of genetic inbreeding may point to continued migration from the steppes after the initial arrival of the Avars in the Carpathian Basin. Alternatively, it may reflect admixture from lower-status groups, not included in this study, who are attested as having arrived together with the Avar elite. The high levels of eastern ancestry among the elite maintained a century and more after the initial immigration may indicate a considerable size of the incoming population and marital networks restricted to it but may also point to later movements and continuing contacts with Central Asia. Both males (N = 15) and females (N = 12) were detected with eastern profiles, which suggests that both women and men had arrived from eastern Asia.

In contrast to the overall uniform northern East Asian genetic profiles of the DTI elites, the individuals retrieved from other early Avar-period elite sites in the neighboring regions are much more heterogeneous. The high-status individuals from the Transtisza group still carry the highest proportions of northern East Asian ancestry found outside the DTI. The more complex admixture patterns (more inter-individual variability in terms of admixing sources, proportions, and dates of admixture) could be the results of different time-layers of connections with the eastern steppe, possibly going back to the Hunnic period or later movements from the steppes to the Carpathian Basin.

Particularly remarkable are the two elite individuals from Kölked in southern Transdanubia, who display a similar western/southern European cultural habitus. However, they show very different genetic histories between each other and with respect to the DTI elites. The woman carries a local unadmixed pre-Avar-period genetic profile and is buried with grave goods associated with late-antique, Merovingian, and Byzantine traditions (Vida, 2018).

The male's genetic ancestry points to the steppes north of the Caucasus or other Iranianate regions.

These results indicate the emergence of a genetically heterogeneous local elite stratum under the rule of the immigrant Avar elite population. Similarly, the more western genetic profile of non-elite early- and late-Avar-period individuals, as far as represented in the present study, shows a stronger connection with the pre-Avar-period population of the Carpathian Basin. A number of recently admixed non-elite individuals, especially in the later period, reveal a variable amount of mostly unidirectional gene flow from the Northeast Asian immigrants to the local population. The Avar Empire was among the longest lasting of a series of political shifts and population changes in the Carpathian Basin between the 5th and 10th centuries (Bálint, 2019; Pohl, 2018). Evidence of similar population shifts during the Hunnic (5th century), Longobard (6th century), or Magyar (9th–10th centuries) migrations is much less robust than the evidence presented here. The only two Hunnic-period genomes available, analyzed above (Hun_P_Budapest_5c and Hun_P_NTransdanubia_5c), suggest a wide genetic variation for this mobile group (see Figure 2A). Data collected from the Longobard period analyzed in Amorim et al. (2018) indicate a heterogeneous population organized along a north-south European cline and, given the lack of genomic data from near-contemporary northern Europe, no clear evidence of a direct connection between the Danubian region and Scandinavia can be drawn. It remains to be seen to what extent genomic data will allow clear distinctions between central/northern European ancestries, for instance, in the Longobard kingdom. For the Magyars, only uniparental markers have been analyzed so far, and are of limited use with regard to the ca. 20%–30% of "central-inner" Asian origin of maternal and paternal lineages (Neparáczki et al., 2018, 2019).

In contrast to these poorly documented periods, our results provide robust genetic support for the Northeast Asian ancestry of the Avar-period elite in the core region of the Avar empire (DTI in the Carpathian Basin) from the middle third of the 7th CE to the early 8th CE (early to middle Avar period). These results also allow us to confidently conclude that an outstanding genetic variability existed in the early Medieval Carpathian Basin, covering almost the entire genetic variation of present-day Eurasia and providing clear evidence of long-distance *trans*-Eurasian migrations.

Limitations of the study

The results in this study indicate the emergence of a genetically heterogeneous local elite stratum under the rule of the immigrant Avar elite population. These results are biased by non-random selection of the burials that we analyzed, and an important direction for future work is to carry out larger-sample-size studies including of entire cemeteries in order to capture, as best as it can be done for past populations, the full spectrum of society. Additional samples and hence resolution from Northeast Asia would also be able to better characterize the source region of the incoming population.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2022.03.007.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 856453 ERC-2019-SyG), the Max Planck Society, the Hungarian Scientific Research Fund (OTKA-NKFI; grant number: NN113157), the Szilágyi Family Foundation, the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (to A.S.-N.), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1C1C1003879), the Allen Discovery Center program (a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation), the Czech Grant Agency (GACR 21-17092X), John Templeton Foundation grant 61220, and the Howard Hughes Medical Institute. We would like to thank Ronny Barr of the Multimedia Department of the Max Planck Institute for Evolutionary Anthropology for the graphical support.

AUTHOR CONTRIBUTIONS

Conceptualization, J.K., C.J., T.V., A.S.-N., G.A.G.-R., D.R., I.K., G.C., and Z.R.; formal analysis, G.A.G.-R., C.J., A.S.-N., and A.B.R.; investigation, A.S.-N., G.B., N.R., V.C., O.C., and B.S.; resources, T.Á.R., A.B., Z.B., N.B., S.C., J.D., Z.F., T. Hága, T. Hajdu, M.J., V.K., B.K., P.M., A.M., B.N.K., C.B., G.M.L., J.G.Ó., M.S., T.S., J.T., Z.T., E.K.T., B.G.M., and R.P.; visualization, G.A.G.-R., C.J., A.S.-N., and A.B.R.; data curation, G.A.G.-R., A.S.-N., and J.K.; writing – original draft, G.A.G.-R., C.J., A.S.-N., I.K., G.C., Z.R., and Z.H.; writing – review & editing, W.P., P.G., T.V., D.R., and J.K.; supervision, G.A.G.-R., Z.H., C.J., and J.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 4, 2021
Revised: January 28, 2022
Accepted: March 4, 2022
Published: April 1, 2022

REFERENCES

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

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.

Alves, I., Arenas, M., Currat, M., Sramkova Hanulova, A., Sousa, V.C., Ray, N., and Excoffier, L. (2016). Long-distance dispersal shaped patterns of human genetic diversity in Eurasia. Mol. Biol. Evol. *33*, 946–958.

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*, 3547.

Bálint, C. (2019). The Avars, Byzantium and Italy: A Study in Chronology and Cultural History (Archaeolingua).

Breuer, E. (2005). Byzanz an der Donau: eine Einführung in Chronologie und Fundmaterial zur Archäologie im Frühmittelalter im mittleren Donauraum (Breuer, Eric).

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

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*, 948.

Curta, F. (2021). The long sixth century in eastern Europe (East Central and Eastern Europe in the Middle Ages, 450–1450) (BRILL).

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

Daim, F. (1992). Awarenforschungen I-II. Archaeologia Austriaca Monographien (Institut für Ur- und Frühgeschichte der Universität Wien).

Daim, F. (2003). Avars and Avar archaeology. An introduction. In Regna, and Gentes: The Relationship Between Late Antique and Early Medieval Peoples and Kingdoms in the Transformation of the Roman World, W. Pohl, H.-W. Goetz, and J. Jarnut, eds. (BRILL), pp. 463–570.

Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, 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.

de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, 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.

Delaneau, O., Howie, B., Cox, A.J., Zagury, J.-F., and Marchini, J. (2013). Haplotype estimation using sequencing reads. Am. J. Hum. Genet. *93*, 687–696.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Dobrovits, M. (2003). "They Called Themselves Avar"—Considering the Pseudo-Avar Question in the Work of Theophylaktos. In Eran ud Aneran. Studies presented to Boris Il'ic Marsak on the occasion of his 70th birthday, M. Compareti, P. Raffetta, and G. Scarcia, eds. (Libreria Editrice Cafoscarina), pp. 175–186.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013b). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227.

Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013a). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

Gansauge, M.-T., Aximu-Petri, A., Nagel, S., and Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. Nat. Protoc. *15*, 2279–2300.

Garam, É. (2005). Avar kori nemzetségfő sírja Maglódon. Das awarenzeitliche Sippenhäuptlingsgrab von Maglód. ComArchHung, 407–436.

Gnecchi-Ruscone, 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*, eabe4414.

Golden, P.B. (2013). Some notes on the Avars and Rouran. In The Steppe Lands and the World Beyond Them : Studies in Honor of Victor Spinei on His 70th Birthday, V. Spinei, F. Curta, B.-P. Maleon, W. Treadgold, and J. Shepard, eds. (Editura Universităţii "Alexandru Ioan Cuza" din Iaşi), pp. 43–66.

Gulyás, B. (2016). One people, two regions? Thoughts on the early Avar period system of relationships in eastern Europe beyond the Tisza river. Hungarian Archaeology. http://www.hungarianarchaeology.hu/?page_id=279#post-6843.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

Hurbanič, M. (2019). The Avar Siege of Constantinople in 626 (Springer).

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.

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.

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.

Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A.S., Taylor, W.T.T., Hagan, R.W., Bromage, S., Tsolmon, S., Trachsel, C., Grossmann, J., et al. (2018). Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. Proc. Natl. Acad. Sci. USA *115*, E11248–E11255.

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.

Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.

Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., et al. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nat. Commun. *8*, 14115.

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., and Kronenberg, F. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. *32*, 25–32.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.

Kradin, N.N. (2005). From tribal confederation to empire: the evolution of the Rouran society. Acta Orient. Acad. Sci. Hung. *58*, 149–169.

Krzewińska, M., Kılınç, 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.

Lamnidis, T.C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., et al. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat. Commun. *9*, 5018.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

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.

Li, J., Zhang, Y., Zhao, Y., Chen, Y., Ochir, A., Sarenbilige, 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.

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

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.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. Genetics *193*, 1233–1254.

Lőrinczy, G. (2017). Frühawarenzeitliche Bestattungßitten im Gebiet der Großen Ungarischen Tiefebene östlich der Theiß. Archäologische Angaben und Bemerkungen zur Geschichte der Region im 6. und 7. Jahrhundert. Acta Archaeol. Acad. Sci. *68*, 137–169.

Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J.V., van Driem, G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C., et al. (2018). The prehistoric peopling of Southeast Asia. Science *361*, 88–92.

Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS One *13*, e0195491.

Nagy, M. (2010). A hun-age burial with male skeleton and horse bones found in Budapest. In Neglected Barbarians. Studies in the Early Middle Ages, F. Curta, ed. (Brepols Publishers), pp. 137–175.

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*.

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.

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.

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*, 2700.

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.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.

Pohl, W. (2018). The Avars: A Steppe Empire in Central Europe (Cornell University Press), pp. 567–822.

R Development Core Team (2021). R: A language and environment for statistical computing : reference index (R Foundation for Statistical Computing).

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

Ralf, A., González, D.M., Zhong, K., and Kayser, M. (2018). Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. Mol. Biol. Evol. *35*, 1820.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374.

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.

Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nat. Commun. *12*, 5425.

Rohland, N., Glocke, I., Aximu-Petri, A., and Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. *13*, 2447–2461.

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.

Salter-Townshend, M., and Myers, S. (2019). Fine-scale inference of ancestry segments without prior knowledge of admixing groups. Genetics *212*, 869–889.

Savelyev, A., and Jeong, C. (2020). Early nomads of the Eastern Steppe and their tentative connections in the West. Evol. Hum. Sci. *2*, e20.

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

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.

Spiliopoulou, A., Colombo, M., Orchard, P., Agakov, F., and McKeigue, P. (2017). GeneImp: fast imputation to large reference panels using genotype likelihoods from ultralow coverage sequencing. Genetics *206*, 91–104.

Szádeczky-Kardoss, S. (1990). The Avars. In The Cambridge History of Early Inner Asia, D. Sinor, ed. (Cambridge University Press), pp. 206–228.

Szenthe, G. (2019). The "Late Avar" Reform and the "Long eighth century": A Tale of the Hesitation between Structural Transformation and the Persistent

Nomadic Traditions (7th to 9th century AD). Acta Archaeol. Acad. Sci. Hung. *70*, 215–250.

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.

Venables, W.N., and Ripley, B.D. (2003). Modern Applied Statistics with S (Springer Science and Business Media).

Vida, T. (2008). Conflict and coexistence: the local population of the Carpathian Basin under Avar rule (sixth to seventh century). In The other Europe in the Middle Ages. Avars, Bulgars, Khazars, and Cumans, F. Curta, ed. (BRILL), pp. 13–46.

Vida, T. (2016). "They asked to Be Settled in Pannonia…" A Study on Integration and Acculturation – the Case of the Avars. In Between Byzantium and the Steppe: archaeological and Historical Studies in Honour of Csanad Balint on the Occasion of His 70th Birthday, Á. Bollók, G. Csiky, and T. Vida, eds. (Institute of Archaeology, Research Centre for the Humanities, Hungarian Academy of Sciences), pp. 51–70.

Vida, T. (2018). Being Avar! A case study for changes in the social display of identity in the early Avar period. In Lebenswelten zwischen Archäologie und Geschichte: Festschrift für Falko Daim zu seinem 65. Geburtstag, J. Drauschke, E. Kislinger, K. Kühtreiber, T. Kühtreiber, G. Scharrer-Liška, and T. Vida, eds. (Monographien des Römisch-Germanischen Zentralmuseums), pp. 419–436.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ancient skeletal element	This study	A1801
Ancient skeletal element	This study	A1802
Ancient skeletal element	This study	A1803
Ancient skeletal element	This study	A1804
Ancient skeletal element	This study	A1805
Ancient skeletal element	This study	A1806
Ancient skeletal element	This study	A1807
Ancient skeletal element	This study	A1808
Ancient skeletal element	This study	A1809
Ancient skeletal element	This study	A1810
Ancient skeletal element	This study	A1811
Ancient skeletal element	This study	A1812
Ancient skeletal element	This study	A1813
Ancient skeletal element	This study	A1814
Ancient skeletal element	This study	A1815
Ancient skeletal element	This study	A1816
Ancient skeletal element	This study	A1817
Ancient skeletal element	This study	A1818
Ancient skeletal element	This study	A1819
Ancient skeletal element	This study	A1820
Ancient skeletal element	This study	A1821
Ancient skeletal element	This study	A1822
Ancient skeletal element	This study	A1823
Ancient skeletal element	This study	A1824
Ancient skeletal element	This study	A1825
Ancient skeletal element	This study	I20801
Ancient skeletal element	This study	I20800
Ancient skeletal element	This study	I20798
Ancient skeletal element	This study	I20799
Ancient skeletal element	This study	I16812
Ancient skeletal element	This study	I16741
Ancient skeletal element	This study	I18742
Ancient skeletal element	This study	I18743
Ancient skeletal element	This study	I18744
Ancient skeletal element	This study	I18224
Ancient skeletal element	This study	I18223
Ancient skeletal element	This study	I18225
Ancient skeletal element	This study	I18222
Ancient skeletal element	This study	I16759
Ancient skeletal element	This study	I18174
Ancient skeletal element	This study	I18184
Ancient skeletal element	This study	I18185
Ancient skeletal element	This study	I16743

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ancient skeletal element	This study	I16744
Ancient skeletal element	This study	I16753
Ancient skeletal element	This study	I16752
Ancient skeletal element	This study	I16751
Ancient skeletal element	This study	I16750
Ancient skeletal element	This study	A181013
Ancient skeletal element	This study	A181014
Ancient skeletal element	This study	A181015
Ancient skeletal element	This study	A181016
Ancient skeletal element	This study	A181017
Ancient skeletal element	This study	A181018
Ancient skeletal element	This study	A181019
Ancient skeletal element	This study	A181020
Ancient skeletal element	This study	A181021
Ancient skeletal element	This study	A181022
Ancient skeletal element	This study	A181023
Ancient skeletal element	This study	A181024
Ancient skeletal element	This study	A181025
Ancient skeletal element	This study	A181026
Ancient skeletal element	This study	A181027
Ancient skeletal element	This study	A181028
Ancient skeletal element	This study	I20802
Ancient skeletal element	This study	A181029
Chemicals, peptides, and recombinant proteins		
Destilled Water DNA free, UltraPure™	Thermo Fisher Scientific	Cat# 10977035
0.5 M EDTA pH 8.0	Thermo Fisher Scientific	Cat# AM9261
Proteinase K	Thermo Fisher Scientific	Cat# AM2548
Isopropanol	Sigma Aldrich	Cat# I9516
Guanidine hydrochloride	Sigma Aldrich	Cat# G4505
Sodium Acetate Solution (3 M), pH 5.2	Thermo Fisher Scientific	Cat# R1181
Tween-20	Sigma Aldrich	Cat# P2287
Buffer PE	Qiagen	Cat# 19065
Buffer PB	Qiagen	Cat# 19066
Tris-EDTA buffer solution	Sigma Aldrich	Cat# 93283
10x Buffer Tango	Thermo Fisher Scientific	Cat# BY5
ATP 100 mM	Thermo Fisher Scientific	Cat# R0441
BSA 20mg/mL	Roche	Cat# 10711454001
dNTP Mix	Thermo Fisher Scientific	Cat# R1121
USER enzyme	New England Biolabs	Cat# M5505
Uracil Glycosylase inhibitor (UGI)	New England Biolabs	Cat# M0281
T4 Polynucleotide Kinase	New England Biolabs	Cat# M0201
T4 DNA Polymerase	New England Biolabs	Cat# M0203
Bst DNA Polymerase, large fragment	New England Biolabs	Cat# M0275L
Ethanol	Merck	Cat# 1009831000
10x T4 Ligase Buffer	Thermo Fisher Scientific	Cat# EL0011
T4 DNA Ligase	Thermo Fisher Scientific	Cat# EL0011
10x Thermopol Buffer	New England Biolabs	Cat# B9004S
Ampure XP	Bioscience	Cat# BCI-A63881

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Agilent D1000 ScreenTapes	Agilent Technologies	Cat# 5067-5582
Agilent D1000 Ladder	Agilent Technologies	Cat# 5067-5586
Agilent D1000 Reagents	Agilent Technologies	Cat# 5067-5583
Agarose	Lonza	Cat# 50004
HyperLadder™ 25bp (formerly HyperLadder V),	Bioline	Cat# BIO-33057
ECO Safe Nucleic Acid Staining Solution 20,000X	Thermo Fisher Scientific	Cat# 3910001
2X Hi-RPM Hybridization Buffer	Agilent Technologies	Cat# 5190-0403
PfuTurbo Cx Hotstart DNA Polymerase	Agilent Technologies	Cat# 600412
Herculase II Fusion DNA Polymerase	Agilent Technologies	Cat# 600679
Sodiumhydroxide Pellets	Fisher Scientific	Cat# 10306200
Sera-Mag Magnetic Speed-beads. Carboxylate-Modified (1 mm, 3EDAC/PA5)	GE LifeScience	Cat# 65152105050250
Dynabeads MyOne Streptavidin	Thermo Fisher Scientific	Cat# 65602
SSC Buffer (20x)	Thermo Fisher Scientific	Cat# AM9770
GeneAmp 10x PCR Gold Buffer	Thermo Fisher Scientific	Cat# 4379874
Salmon sperm DNA	Thermo Fisher Scientific	Cat# 15632-011
Human Cot-I DNA	Thermo Fisher Scientific	Cat#15279011
5M NaCl	Sigma Aldrich	Cat# S5150
1M NaOH	Sigma Aldrich	Cat# 71463
1 M Tris-HCl pH 8.0	Sigma Aldrich	Cat# AM9856
50x Denhardt's solution	Thermo Fisher Scientific	Cat# 750018
Methanol, certified ACS	VWR	Cat# EM-MX0485-3
Acetone, certified ACS	VWR	Cat# BDH1101-4LP
Dichloromethane, certified ACS	VWR	Cat# EMD-DX0835-3
Hydrochloric acid, 6N, 0.5N & 0.01N	VWR	Cat# EMD-HX0603-3
5 M Sodium chloride solution	Sigma-Aldrich	Cat# S5150-1L
20% SDS	Serva	Cat# 39575.01
PEG-4000	Thermo Fisher Scientific	Cat# EL0011
PEG-8000	Promega	Cat# V3011
Critical commercial assays		
MinElute PCR Purification Kit	QIAGEN	Cat# 28006
TwistAmp Basic Kit	TwistDX	Cat# TABAS03kit
Qubit® dsDNA HS Assay Kit, 500 assays	Thermo Fisher Scientific	Cat# Q32854
High Pure Extender Assembly from the Roche High Pure Viral Nucleic Acid Large Volume Kit,40 reactions	Roche	Cat# 5114403001
MiSeq Reagent Kit v3 (150 cycle)	Illumina	Cat# MS-102-3001
NextSeq® 500/550 High Output Kit v2 (150 cycles)	Illumina	Cat# FC-404-2002
HiSeq Cluster Kit SR	Illumina	GD-410-1001
HiSeq 4000 SBS Kit (50/75 cycles)	Illumina	Cat# FC-410-1001/2
NextSeq® 500/550 High Output Kit v2 (150 cycles)	Illumina	Cat# FC-404-2002
DyNAmo Flash SYBR Green qPCR Kit	Thermo Fisher Scientific	Cat# F415L
Maxima SYBR Green kit	Thermo Fisher Scientific	Cat# K0251
Oligo aCGH/Chip-on-Chip Hybridization Kit	Agilent Technologies	Cat# 5188-5220

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data (European nucleotide archive)	This study	ENA: PRJEB50368
1240K Genotype data (Edmond Data Repository of Max Planck Society)	This study	https://edmond.mpdl.mpg.de/
Software and algorithms		
EAGER 1.92.55	Peltzer et al., 2016	https://eager.readthedocs.io/en/latest/
AdapterRemoval 2.2.0	Schubert et al., 2016	https://github.com/MikkelSchubert/adapterremoval
BWA 0.7.12	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
DeDup 0.12.2	Peltzer et al., 2016	https://github.com/apeltzer/DeDup
mapDamage 2.0.6	Jónsson et al., 2013	https://github.com/ginolhac/mapDamage
bamUtil 1.0.13	https://github.com/statgen/bamUtil	https://github.com/statgen/bamUtil
CircularMapper	Peltzer et al., 2016	https://github.com/apeltzer/CircularMapper
ANGSD 0.910	Korneliussen et al., 2014	http://www.popgen.dk/angsd/index.php/ANGSD
Schmutzi	Renaud et al., 2015	https://github.com/grenaud/schmutzi
SAMtools 1.3	Li et al., 2009	http://www.htslib.org/doc/samtools.html
pileupCaller	https://github.com/stschiff/sequenceTools	https://github.com/stschiff/sequenceTools
GATK v3.5	DePristo et al., 2011	https://gatk.broadinstitute.org/hc/en-us
GeneImp 1.4	Spiliopoulou et al., 2017	https://pm2.phs.ed.ac.uk/geneimp/
SHAPEIT v2.r790	Delaneau et al., 2013	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
pMMRCalculator	https://github.com/TCLamnidis/pMMRCalculator	https://github.com/TCLamnidis/pMMRCalculator
HaploGrep2	Kloss-Brandstätter et al., 2011	https://haplogrep.i-med.ac.at/
Yleaf v1.0	Ralf et al., 2018	https://github.com/genid/Yleaf
READ	Monroy Kuhn et al., 2018	https://bitbucket.org/tguenther/read/src/master/
lcMLkin	Lipatov et al., 2015	https://github.com/COMBINE-lab/maximum-likelihood-relatedness-estimation
EIGENSOFT v6.0.1	Patterson et al., 2006	https://github.com/DReichLab/EIG
ADMIXTOOLS 5.1	Patterson et al., 2012	https://github.com/DReichLab/AdmixTools
DATES v753	Narasimhan et al., 2019	https://github.com/priyamoorjani/DATES
MOSAIC v1.3	Salter-Townshend and Myers, 2019	https://maths.ucd.ie/~mst/MOSAIC/
RFMix v2.03	Maples et al., 2013	https://github.com/slowkoni/rfmix
PLINK v. 1.9	Chang et al., 2015	https://www.cog-genomics.org/plink/
hapROH v0.3	Ringbauer et al., 2021	https://pypi.org/project/hapROH/
R v4.0.5	R Development Core Team, 2021	https://www.r-project.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Johannes Krause (krause@eva.mpg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The newly produced aligned sequence data are deposited in the European Nucleotide Archive (ENA) with the following accession number: PRJEB50368. Haploid genotype data for the 1240K panel is available in eigenstrat format via the Edmond Data Repository of the Max Planck Society (https://edmond.mpdl.mpg.de/).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological material and ethics permission

The authors declare that they had requested and got permission from the stakeholders, excavator and processor anthropologists and archeologists for the destructive DNA analyses of the anthropological material presented in this study.

We generated new genome-wide data from skeletal remains of 66 ancient individuals from the following sites (see Data S1 for archeological details):

Sarmatian period sites

Kecskemét, Mindszenti-dűlő (n = 8); Archaeological context newly reported.
Hajdúnánás, Fürj-halom-dűlő (n = 8); Archaeological context newly reported.
Derecske, Karakas-dűlő (n = 1); Archaeological context newly reported.

Hun period sites

Árpás, Dombiföld-Szérűskert (n = 1); Archaeological context newly reported.

Avar period sites

Békésszentandrás, Benda-tanya (n = 1); Archaeological context reported in Csáky et al. (2020).
Berettyóújfalu, Nagy-Bócs-dűlő (n = 1); Archaeological context newly reported.
Budapest, Csepel-Kavicsbánya (n =1); Archaeological context reported in Csáky et al. (2020).
Kunpeszér, Felsőpeszéri út (n = 6); Archaeological context reported in Csáky et al. (2020).
Kecskemét, Sallai út (n = 1); Archaeological context reported in Csáky et al. (2020).
Debrecen, Bordás-tanya (n = 1); Archaeological context newly reported.
Derecske, Bikás-dűlő (n = 1); Archaeological context newly reported.
Derecske, Hosszú-lapos (n = 2); Archaeological context newly reported.
Derecske, Karakas-dűlő (n = 2); Archaeological context newly reported.
Hajdúböszörmény, Homokbánya IV (n = 1); Archaeological context newly reported.
Kölked, Feketekapu (n = 2); Archaeological context newly reported.
Kövegy, Nagy-földek (n = 1); Archaeological context newly reported.
Kunbábony (n = 1); Archaeological context reported in Csáky et al. (2020).
Kunszállás, Fülöpjakab (n = 7); Archaeological context reported in Csáky et al. (2020).
Petőfiszállás (n = 1); Archaeological context reported in Csáky et al. (2020).
Albertirsa Site 22 (n = 2); Archaeological context newly reported.
Albertirsa, Szentmártoni út (n = 5); Archaeological context newly reported.
Alsónyék, Elkerülő út (n = 1); Archaeological context newly reported.
Szalkszentmárton, Táborállás (n = 1); Archaeological context reported in Csáky et al. (2020).
Szarvas, Kovács-halom (n =4); Archaeological context reported in Csáky et al. (2020).
Tiszapüspöki, Holt-Tisza-part and Felső-földek (n = 3); Archaeological context newly reported.
Visonta, Nagycsapás (n = 3); Archaeological context newly reported.

Historical background

In the 5th and 6th centuries, the Carpathian Basin (modern Hungary and the adjacent lowlands of southern Slovakia, western Romania, northern Serbia and eastern Croatia and Austria) was a region particularly affected by migrations and changes of political dominion, which are well-documented in contemporary Latin and Greek texts. Since the beginning of the first millennium, the Roman Empire had ruled over the land west and south of the Danube, most of it in the province of Pannonia. Sarmatians, who had arrived from the Pontic Steppes north of the Black Sea in the 1st century CE, lived between the Danube and the mountains of Transylvania. Around 400, this relatively stable configuration was shattered by the arrival of the Huns. These Huns had arrived from the Central Asian steppes in the lands north of the Black Sea in c. 375. In this region, the Goths, a people of Germanic language, had ruled for some time; now, part of them became subjects of the Huns, while others flew to Roman territory.

The arrival of the Huns toppled the already precarious balance between the Roman Empire and its 'barbarian' neighbours (modern scholars use this contemporary derogatory term for people outside the Ancient Civilisation in a descriptive sense, for lack of a better word). In the course of what is often called 'the Great Migration', the Western Roman Empire dissolved, until the last western emperor was dethroned in a coup by his army in 476. Many scholars have regarded this as a causal relation: barbarians caused 'the Fall of Rome'. However, internal problems of Roman society and the instability of its political system rather facilitated the barbarian takeover than only being its consequences. The increasing Roman demand for soldiers was a decisive pull factor for barbarian immigrants.

Also, the Eastern Roman Empire that we call Byzantium continued to exist for another millennium. Still, barbarian raids caused much damage, and in the course of events, the Roman order was destroyed in some frontier provinces, such as Pannonia.

The contemporary sources classify the many barbarian groups increasingly operating on Roman territory by ethnic terms. This was certainly not inaccurate, because we also get some glimpses of self-identification with these peoples, Huns, Goths, Franks or Longobards, in the sources. Ethnicity was supposed to be based on common origin, but ethnic categories were in reality handled rather flexibly. Changes of ethnic affiliation are repeatedly reported, and the composition of migrating groups often shifted. For instance, the Roman diplomat Priscus met a Hun warrior at Attila's court, who turned out to have been a Roman prisoner of war who now preferred the life of the barbarians. Thus, we should not take these ethnonyms for granted as indicators of actual common ancestry.

In the Carpathian Basin, the initially dispersed groups of Hunnic steppe riders united and initiated a series of disastrous raids on the Balkan provinces, Gaul and Northern Italy. In circa 450 CE, under the rule of Attila, the short-lived Empire of the Huns reached the apogee of its power. After the death of Attila in 453, the Hunnic realm collapsed, and its numerous components (in majority, Germanic-speaking groups) built their own realms, or joined the Roman forces. Along the margins of the Carpathian Basin, Rugians, Heruls, Sciri, Suebi and Sarmatians created ephemeral regional kingdoms which were all gone by 500. The only relatively stable kingdom that emerged was that of the Gepids along the Tisza river, which lasted until 567. In Pannonia, Ostrogoths established their rule, but then crossed into Roman heartlands in the 470s and ultimately conquered Italy in the 490s (where their kingdom lasted until 552/54, for some time expanding back into Pannonia). At that point, Longobards began to settle in Pannonia. In the mid-6th century, they came into conflict with the Gepids, whose kingdom they finally destroyed in 567.

A year later, the Longobards collected their forces and those of numerous other groups (Gepids, Suebi, Pannonian provincials, Sarmatians and others), and marched into Italy, where the kingdom they established played a major role until the Frankish king Charlemagne conquered it in 774. Unlike in the Carpathian Basin, a late-antique society had more or less subsisted in Italy, with urban centers, a well-organised Christian Church and legally defined property rights over a dependent class of farmers. Only remnants of late-antique culture had survived in Pannonia into the 6th century, and with few exceptions the Roman cities were largely deserted. The Carpathian Basin was rather thinly populated in most parts, except for the Gepid core areas. Its population seems to have been rather mixed in origin, although the dominant warrior groups shared many forms of cultural expression – among them, inhumation of the deceased with full dress and grave goods, which allows us to obtain precious information about some aspects of their lifestyle and funerary practices.

When the Gepids were crushed and the Longobards left, the Avars took advantage of the situation and moved into the Carpathian Basin, which they would dominate for over 200 years. They had arrived in the Pontic Steppes north of the Black Sea about ten years before, led by the khagan (which was their rulers' title) Baian. Soon, they defeated the mostly smaller groups of steppe warriors who had replaced the Huns in the course of the later 5th century. These were mostly Turkic-speaking peoples such as the Utigurs, the Cutrigurs or the Onogurs, sometimes generically called Bulgars in the texts. North of the Lower Danube, the Avars encountered Slavic populations who had only appeared recently in the written record. The Avars had entered into an alliance with the Byzantine (or East Roman) Empire in 558/59, and often sent envoys to its capital Constantinople (modern Istanbul). Only after they had consolidated their rule over the Carpathian Basin did they start a series of major raids into the Balkan provinces, culminating in a siege of Constantinople, together with a Persian army, in 626. The siege failed, and the mobilizing power of the Avar khagans collapsed. Unlike the Huns, they rarely attacked their western neighbors, the Longobards in Italy and the Franks, whose Merovingian dynasty dominated much of modern France and Germany.

Like the Huns, the Avars had arrived from the Central Asian steppes. We have better information about the circumstances of their migration from written sources than about the Huns. With the Huns, we can only assume that they were in some ways connected to the Xiongnu, who had ruled over a powerful steppe empire centered in modern Mongolia and repeatedly challenged the Chinese empire of the Han Dynasty between the 2nd century BCE and the 1st century CE. Further west, Huns only appear in the 4th century: the Hephthalites and some smaller realms in the steppes between the Caspian Sea and the Hindukush, and, of course, the European Huns. In the east, Xiongnu power was eventually replaced by two other steppe peoples. One of them were the Xianbei, who mostly lived in what is now Manchuria, temporarily extending their rule over parts of Mongolia in the 2nd/3rd century and over Northern China in the 5th/6th centuries. The other people were called Rouran by the Chinese, and ruled by khagans. It is quite likely that they were known as Avars in the steppe. Their empire was destroyed by the emerging khaganate of the Turks in 552.

What happened then is described in reports based on diplomatic exchanges between the Turks and Byzantium. The Turks denied that the European Avars were directly descended from the Rouran, which would have given their khagans the ancient legitimacy of the Rouran khaganate. They claimed that those who had fled Turkish expansion in the 550s were a mixed people called Warchonites, composed of groups of former Rouran subjects, mainly Ogurs who had lived in western Central Asia. They had only adopted the prestigious name Avars to frighten those whom they encountered on their westward migration. Indeed, we know from Chinese sources that many among the Rouran had been killed by Turks and Chinese, and others had fled eastward, and reputedly ended up in Korea. Still, it is not unlikely that we can assume that groups from the Western Eurasian steppes took part in the Avar migration, and that at least the Avar khagans and their core group were actually descended from the Rouran. The results of the present article support this view. The written sources also indicate that Cutrigurs, Utigurs and Bulgars from the Pontic steppes accompanied the Avars into the Carpathian Basin.

In the early Avar Empire, the archaeological record demonstrates that the population of the Carpathian Basin remained quite heterogeneous, at least culturally. The late-Roman element was reinforced by captives from the Balkan provinces and elsewhere settled

as dependent laborers. In the course of the 7th century, following the unsuccessful siege of Constantinople, the cultural differences within the Avar settlement area largely faded out. Whether or not this led to wide-spread population admixture remains to be seen and is one of the primary issues being investigated through HistoGenes, a European Research Council (ERC) project analysing genomic, historical, and archaeological data in the Carpathian Basis between the fifth and tenth centuries.

Archaeological background

The Avars arrived to the Carpathian Basin at the end of the 6th century and united local peoples and communities of the Carpathian Basin, including the remnants of the Romanised population in the territory of former Pannonia and various Barbarian groups with heterogeneous background (Daim 2003; Bálint 2019). This heterogeneity is attested in its rich archaeological material consisting of ca. 600 settlements and approx. 100,000 excavated burials. After its peak in the mid-7th century the Khaganate remained a regional political power until the 8th century. The Avar Khaganate was made up of networks of overlapping groups organized along territorial, cultural, social and other lines.

Concentration of high-status male burials (Bócsa-Kunbábony circle) suggests that the 7th-century Avar power center lay in the DTI. These members of the early Avar elite (the leaders of the early Avar polity and the khagan's military retinue) were buried with gold- and silver-decorated weapons and belts, various insignias and valuable prestige objects. These artefacts are superbly crafted products of the period's goldsmithing that indicate their owners' prominent social position and their long-distance connections and that they remained part of the network that is the Eurasian steppe. Nomadic insignia of rank and power could have been an eagle-headed gold sheet, it conjectured that it had perhaps adorned the sceptre. The lunula from a Kunbábony burial fits into an Inner Asian aristocratic tradition (Vida 2016; Bálint 2019).

A group with different material culture, interpreted as Eastern European nomads settled the Trans-Tisza region. Their burials share numerous similarities with 6th-7th-century Eastern European population and include Byzantine-influenced jewellery items and dress accessories probably adopted in the Pontic region. These groups are characterised by solitary burials or small burial grounds in the Early Avar period, the deposition of partial or complete sacrificial animals, and ledge or niche graves (Gulyás 2016; Lőrinczy 2017). Some elements of these burial customs can be traced until the end of the Avar period.

The West-Carpathian Basin (the former Roman province of Pannonia) served as border zone of Khaganate, where archaeological material attests the survival and coexistence of various local groups with heterogeneous cultural background showing strong connection to the Mediterranean (certain brooch and earring types) as well as the Merovingian world in Central and Western Europe. The archaeological heritage of the surviving local communities showing western European connections comprises burials with weapons (spatha, spear, sax) and Merovingian dress accessories (suspended belts, leg bindings, etc.). The rich, high quality find assemblages in this region (for example Kölked) probably reflect a community with internal autonomy, which was obliged to render armed service to the Avars under the leadership of their own local elite (Vida, 2008).

The Avar elite cemented the peoples of the Carpathian Basin with an immense integrative power in the 7th century and ensured the powerful political position of the Avars in Europe. The ''speed'' and success of acculturation and integration is reflected by the emergence of a uniform material culture across the Avar Khaganate within the 8th century, expressing the unity of the Avar political organisation (Szenthe, 2019).

METHOD DETAILS

Archeological dating

Three samples have been ^{14}C dated and in these cases ^{14}C results agree with archeological dating (Data S1). The remaining samples have been dated archaeologically. The core of our work (the elite burials) have a very well established archeological dating based on the types and type chronology of artifacts found in the graves (pseudo-buckles, jewelry) (Bálint, 2019; Breuer, 2005; Daim, 2003; Garam, 2005). Since they contain such a high concentration of fast-changing types, their dating can be considered very accurate within a few decades. The empty or poorly furnished burials are more problematic for dating but constitute a minority of our samples and they were dated based on their larger context.

Ancient DNA processing and sequencing

Laboratory experiments were performed at the Ancient DNA clean room facilities of the Institute of Archaeogenomics, Research Centre for the Humanities, Budapest, at the laboratory of the Max Planck Institute for the Science of Human History, Jena, at the Ancient DNA Lab, University of Vienna, and at Harvard Medical School, Boston. From each bone or tooth specimen, 30 to 70 mg of powder was sampled and used for the following steps. DNA extraction and genomic library preparation of 42 samples labelled with an initial ''A'' were done in Budapest, following well-established ancient DNA workflow protocols (Csáky et al., 2020; Lipson et al., 2017). The DNA extraction was performed based on the protocol of (Dabney et al., 2013) with some modifications described also by (Lipson et al., 2017). DNA libraries were prepared using UDG-half treatment methods (Rohland et al., 2015). We included library negative controls and/or extraction negative controls in every batch. Unique P5 and P7 adapter combinations were used for every library (Rohland et al., 2015). Double barcoded adaptor-ligated libraries were then amplified with TwistAmp Basic (Twist DX Ltd), purified with AMPure XP beads (Agilent) and checked on a 3% agarose gel. Mitogenome capture was performed according

to Csáky et al. (2020) in Budapest and sequenced with 2x75 cycles on Illumina MiSeq. This study reports new mitogenome sequences for the samples A181013-A181029, others from the Budapest lab were published in Csáky et al. (2020).

In Jena, an equimolar pool of 42 UDG-half libraries was prepared for shotgun sequencing on an Illumina HiSeq 4000 platform using a single-end 76-cycle sequencing kit. All of the sequenced libraries showed high human endogenous DNA proportions (Table S1) and ancient DNA characteristic damage patterns at the end of the fragments (5' end at least ~0.05%; Table S1). The libraries were enriched using an in-solution enrichment protocol consisting of DNA probes specifically targeting 1,237,207 single nucleotide markers (SNPs) genome-wide known to be variable in worldwide sets of human populations (Fu et al., 2013b; Haak et al., 2015; Mathieson et al., 2015). For this, all libraries we re-amplified using the Herculase II Fusion DNA Polymerase (Agilent) to achieve 1-2 mg of total DNA in 5.2 ml (200-400 ng/ml), they were then purified using the MinElute DNA purification kit (QIAGEN) and their concentrations were measured on a NanoDrop spectrophotometer (Thermo Fisher Scientific).

In Boston, all 24 samples (between 33 and 42 mg of powder) were incubated in 750 μl lysis buffer overnight and from 1/5th of the lysis buffer DNA was extracted using Dabney binding buffer and silica magnetic beads utilizing an automated workflow on an Agilent Bravo workstation (Rohland et al., 2018). The entire extract was then converted into libraries; 6 of the 24 libraries were built using a single-stranded library preparation protocol following USER treatment and finished with dual indexing (Gansauge et al., 2020), and 18 libraries were built using a double-stranded library preparation that retains a subset of terminal Uracils, but removes all them internally (Rohland et al., 2015); both library preparations were run on Agilent Bravo workstations and the double-stranded preparation utilized silica magnetic beads instead of MinElute columns for the cleanup steps to be compatible with an automated workflow. Target enrichment of above-mentioned nuclear SNPs (and MT genome) was done as referenced above.

Overall, the 66 enriched libraries were sequenced to 0.01-5.3x coverage for the "1240K" target sites (with a median of 2.9x), covering 13,749 to 1,119,583 target SNPs at least once (with a median of 975,215 SNPs).

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequence data processing

The *fastq* files containing the sequenced reads were processed through the EAGER 1.92.55 pipeline wrapper (Peltzer et al., 2016). Prior to running EAGER, the first 7 bps of each read were trimmed to remove internal barcode sequences using a custom python script. To remove adaptors and short reads (< 30 bp) AdapterRemoval v2.2.0 was used (Schubert et al., 2016). The reads were then mapped to the Human Reference Genome Hs37d5 with bwa v0.7.12 aln/samse alignment algorithm (Li and Durbin, 2009) with the parameters "-n" and "-l" set to 0.01 and 32 respectively. The reads with phred mapping quality $\geq$ 30 were then discarded using "-q" (q30-reads) in Samtools v1.3 (Li et al., 2009). We then used DeDup v0.12.2 to remove PCR duplicates (Peltzer et al., 2016). To estimate the amount of C to T taphonomic deamination at the ends of the mapped fragments we used mapDamage v2.0 (Jónsson et al., 2013) run on a subset of 100,000 q30-reads. Exogenous human autosomal DNA contamination was estimated in males by assessing the X chromosome heterozygosity levels with ANGSD v0.910 (Korneliussen et al., 2014) and mitochondrial DNA contamination in males and female was estimated via Schmutzi (Renaud et al., 2015; Table S1). PileupCaller (https://github.com/stschiff/sequenceTools) was used to carry out genotype calling from the q30 reads with the "–randomHaploid" that calls haploid genotypes by randomly choosing one high quality base (phred base quality score $\geq$ 30) on the 1240K panel (pseudodiploid calls). Given that UDG-half treated libraries still preserve a certain amount of C to T deamination at the last 2 bp of the mapped fragments, the transition alleles were called after masking 2 bp at both ends of the q30 reads with the trimBam module of bamUtil v.1.0.13 (Jun et al., 2015) while transversions were called on the un-masked reads.

Sequencing data (bam files) were screened for SNPs collected from the Y chromosomal tree of ISOGG version 15.34 (https://isogg.org/tree/) by Yleaf v1.0 (Ralf et al., 2018), which resulted in Y-haplogroup classification that were checked manually and presented with defining terminal SNPs in Table S1. Sequences from mitogenome capture were aligned with bwa aln applying parameters -o 1 -e 5 -n 0.01 -l 32, after barcode trimming and merging the fastq files. Mapping quality was ensured by samtools view command and samtools rmdup removed the PCR duplicates from the bam files. We used Schmutzi to obtain the consensus sequence of the Mitochondrial DNA and HaploGrep2 (Kloss-Brandstätter et al., 2011) to assign haplogroups (Table S1).

Imputation

The 2-bp masked q30 reads of 66 Avar individuals were used to call genotype likelihoods (GL) with GATK v3.5 (DePristo et al., 2011). GL were called for all the 29,083,171 biallelic positions with a minor allele count 5 or higher contained in the 1000 Genomes Phase 3 release (1KG; 1000 Genomes Project Consortium et al., 2015) using the "UnifiedGenotyper" module with a mean base quality score higher than 30 "-mbq 30". These GL were then used for statistical imputation via GeneImp v1.4 software (Spiliopoulou et al., 2017) by imputing all the 1KG SNPs and using all the 2,504 individuals of the 1KG statistically phased genomes as reference dataset. Imputation was run independently using three different window lengths "kl" {15, 20, 25}, following the developers' instructions. For each imputed SNP the average genotype call probability (GP) across the three different runs was considered and only the SNPs overlapping the 1240K panel were then finally extracted and merged with the 1240KHO dataset. We first qualitatively assessed the performance of imputation considering different posterior probability (PP) thresholds for calling a non-missing genotype (i.e. 0.9, 0.95, 0.99, 0.999). This was done by comparing the results from the imputed and pseudodiploid data on the following analyses: projecting them on the PCA calculated on present-day Eurasian populations, f4-statistics in the form f4(Chimp,

Reference;indX_pseudohap,indX_imputed), runs of homozygosity (Figure S4B) and pairwise mismatch rate (pMMR) using the tool pMMRCalculator (https://github.com/TCLamnidis/pMMRCalculator) that extends the pMMR formula to consider diploid data with heterozygous genotypes that could result in a partial match (i.e. 0.5, if one allele matches but not the other) (Figure S2; Table S1). The 0.9 PP threshold was retained for downstream analyses as it overall gave results more consistent with the respective pseudodiploid data.

To test for the effect of spurious base calls on imputation, we fit a linear regression model with the mean pMMR for each individual as the response variable, and the mean coverage and contamination estimate as predictor variables using the R statistical software package (R Development Core Team, 2021). Since we used contamination estimates from ANGSD, we filter for only male individuals, and remove any closely related pairs. Using the *boxcox* function from the *MASS* package (Venables and Ripley, 2003), we identified that a log transformation was required. Contamination was found to not be a significant predictor for the mean change in pMMR estimates (p=0.152), but coverage was found to be significant (p=7.539e-9).

Evaluation of imputation performance

The pMMR calculated on the pseudodiploid or on the diploid imputed should be equivalent for the same pair of individuals. Hence we developed an algorithm based on the mean difference of the pMMR values to identify an empirical threshold for including or excluding imputed samples, while simultaneously taking into account the difference in the number of overlapping sites due to sample quality and imputation.

Consider N individuals with associated, ordered mean coverage $\boldsymbol{c} = (c_1, \ldots, c_N)$. For individuals j and k let n^h_{jk} and n^i_{jk}, be the number of overlapping sites, and x^h_{jk} and x^i_{jk} be the number of observed mismatches, for the pseudo-haploid and the imputed data respectively. If we assume that these mismatch counts follow a binomial distribution, *i.e.*, $x^h_{jk} \sim \mathrm{B}(n^h_{jk}, p^h_{jk})$ and $x^i_{jk} \sim \mathrm{B}(n^i_{jk}, p^i_{jk})$, then we can calculate a z-score for the mismatch counts of the form

$$z_{jk} = \frac{\left(\widehat{p}^h_{jk} - \widehat{p}^i_{jk}\right)}{\sqrt{\widehat{p}_{jk}\left(1 - \widehat{p}_{jk}\right) \Big/ \left(n^h_{jk} + n^i_{jk}\right)}},$$

where

$$\widehat{p}^h_{jk} = x^h_{jk} \Big/ n^h_{jk}, \widehat{p}^i_{jk} = x^i_{jk} \Big/ n^i_{jk} \text{ and } \widehat{p}_{jk} = \left(x^h_{jk} + x^i_{jk}\right) \Big/ \left(n^h_{jk} + n^i_{jk}\right).$$

We then calculate a mean z-score for all pairwise z-scores of the form

$$\bar{z} = \frac{2}{N(N-1)} \sum_{j,k} z_{jk},$$

with standard deviation σ_z found via 1000 bootstrap samples.

We then find the smallest value of c_m such that if we omit individuals with mean coverage less than or equal to c_m, we observe that

$$\left| \underline{z} \Big/ \sigma_z \right| < 2,$$

indicating that the pMMR values for the pseudo-haploid and the imputed pairs are not significantly different (Figure S2A).

Genetic relatedness

Genetic relatedness between individuals was assessed by measuring the pairwise mismatch rate (pMMR. *i.e.* rate of mismatching alleles) between each pair of individuals (Jeong et al., 2018; Kennett et al., 2017). The pairwise mismatch rate provides an indication of close genetic relationships, such as identical individuals/twins, first and second degree relatives (Jeong et al., 2018; Kennett et al., 2017). We also calculated a similar statistic implemented in READ (Monroy Kuhn et al., 2018) that instead uses a windowed approach that allows to calculate standard errors for the estimated degrees of relatedness, and tests for relationships up to the second degree. Finally, we also employed lcMLkin (Lipatov et al., 2015), a method that instead uses genotype likelihoods to estimate kinship parameters, allowing us to differentiate between parent/child and sibling/sibling relationships for first degree relatives (Figure S4C).

Compilation of population genetic data

The genotype data of 66 individuals was merged with a reference genome-wide panel of 2280 modern individuals genotyped with the microarray technology using the commercial HumanOrigins chip (Jeong et al., 2019; Lazaridis et al., 2016; Patterson et al., 2012) and previously published ancient individuals' genotypes obtained with the same 1240K capture sequencing or a pull down of 1,240K sites from shotgun sequencing data (Allentoft et al., 2015; Amorim et al., 2018; de Barros Damgaard et al., 2018; Damgaard et al., 2018; Fu et al., 2016; Gnecchi-Ruscone et al., 2021; Jeong et al., 2018, 2019, 2020; Krzewińska et al., 2018; Lamnidis et al., 2018; Lazaridis et al., 2016; Li et al., 2018; Mathieson et al., 2015, 2018; McColl et al., 2018; Narasimhan et al., 2019; Ning et al., 2020; Raghavan et al.,

2014; Sikora et al., 2019; Unterländer et al., 2017) downloaded from the Allen Ancient DNA Resource (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data) or merged in-house. We formed two different datasets used for the different analyses as specified below, the "1240K" dataset contaning all the SNPs overlapping the 1,240K panel and the "1240KHO" containg the SNPs overlapping with the HumanOrigins SNP chip (~600K sites).

Population genetic analysis

Principal component analysis (PCA) was calculated with smartpca v16000 in EIGENSOFT v6.0.1 package (Patterson et al., 2006) on the 1240KHO dataset using the lsqproject and the autoshrink options to project the genotypes of the ancient individuals (containing variable amounts of missing data) on top of the principal components calculated on the set of modern populations. PCA was computed on the set of West to East Eurasian populations excluding South Asians of the 1240KHO dataset as well as on a set of West Eurasian only groups (comprehending Continental Europe, the Mediterranean area and the Caucasus only; Figures 2 and S1) for full list of modern individual used see (Gnecchi-Ruscone et al., 2021).

The software *qpAdm* (v632) of the ADMIXTOOLS package (Patterson et al., 2012) was used to run the *f4*-statistic based ancestry analyses on the 1240K dataset (Haak et al., 2015; Reich et al., 2012). Standard errors (SE) for the computed f-statistics were estimated using a block jackknife with 5 centiMorgan (cM) block. We used the default "allsnps: NO" parameter therefore calculating all the underlying *f4*-statistics using the SNP overlap between all groups for each test and excluded three individuals with < 200,000 SNPs covered on the 1240K dataset. Two sets of outgroups were used that included representatives of Eurasian and relevant non-Eurasian ancient lineages when available otherwise present-day proxies. OG1 includes two Paleolithic human branches (Villabruna and MA1): Mbuti.DG, Levant_N, Anatolia_N, Iran_N, Villabruna, Onge.DG, Mixe.DG, DevilsCave_N.SG, MA1, Kolyma_M.SG, YR_LN.

OG2 includes more recent representatives for these lineages (Iron_Gates_HG, EHG,Russia_Bolshoy): Mbuti.DG, Levant_N, Anatolia_N, Iran_N, Iron_Gates_HG, EHG, Onge.DG, Mixe.DG, DevilsCave_N.SG, Russia_Bolshoy, Kolyma_M.SG, YR_LN. We didn't observe major differences between the two sets, showing the robustness of the modelings to certain variation in the outgroups. We did observe cases where OG2 had higher resolution (i.e. higher capacity to distinguish between different combinations of close proxies) because of the ~200,000 SNPs increase in coverage using this outgroup set for each test. We therefore considered OG2 for the individual-based models for drawing historical inferences. qpAdm modeling on the imputed and local ancestry masked data was run on the 1240KHO dataset (as the HO modern populations were used to infer ancestry tracts, only the overlapping sites could be used) and therefore the OG2 outgroup set was slightly modified to include the higher number of individuals available on the HO panel for the modern populations (HO populations without the.DG suffix): Mbuti, Levant_N, Onge, Iran_N, Iron_Gates_HG, EHG, Mixe, Anatolia_N, DevilsCave_N.SG, Russia_Bolshoy, Kolyma_M.SG, YR_LN.

In performing the 2- or 3-way individual-based qpAdm modeling reported in Figure 3A we used the following rationale: 1) if the 2-way models on the pseudodiploid data using AR_Xianbei_P_2c + a combination of local and steppe "western" sources has the resolution to discriminate between local or a steppe source, these results were taken (Figures S1C and S7; Tables S2 and S4). 2) for individuals showing no evident signs of admixture with non-European components, 2-way models contrasting different local sources were tested (Figure S5B). 3) for the DTI elite individuals that showed a higher degree of homogeneity (genetic and archaeological) grouped based analyses were also performed and these higher resolution tests were considered in the final selection (Table S3). 4) for the cases where the grouping would not have been meaningful (e.g. because of the high genetic heterogeneity of the Transtisza group) or when the group-based models did not completely resolve the admixing sources (i.e. both local and steppe sources work for the DTI_late_elite), we performed individual based 3-way models including in the same model the competing local vs steppe sources + AR_Xianbei_P_2c (Figure S1D). For the DTI late case we performed additional group-based tests with two different subgroups showing opposing signal in the 3-way models (DTI_late_elite1 and DTI_late_elite2). We finally verified these results based on 2-way individual-based models using the masked local ancestry data (Figure S2B).

DATES v.753 (https://github.com/priyamoorjani/DATES) was used to estimate the dates the eastern steppe admixture events occurred in the Avar period individuals / populations. This method is based on the same general principle as other admixture dating methods (Loh et al., 2013; Narasimhan et al., 2019), and assuming a 2-way admixture event (i.e. two admixing sources) it captures the decay of ancestry covariance coefficients (AC) which can be calculated between every pair of available SNPs over increasing genetic distance windows (Narasimhan et al., 2019). An exponential function is then fitted to the decay of Weighted AC in order to infer the number of generations since admixture (Loh et al., 2013). As sources we used the Sarmatians as both pre-Avar and pre-Hunnic period local population and as eastern steppe proxy we chose the temporary distal but more importantly un-admixed ANA proxy and well covered LBA/IA group of Ulaanzuukh_SlabGrave in Mongolia (Jeong et al., 2018; Figures 4 and S3).

Haplotype phasing and local ancestry analyses

Haplotype phasing for the imputed ancient data together with the set of worldwide modern populations present in the 1240KHO dataset was performed using SHAPEIT2 v2.r790 (Delaneau et al., 2013) with default parameters and using HapMap phase 3 recombination maps.

Local ancestry analyses were performed using MOSAIC v1.3 (Salter-Townshend and Myers, 2019). For each analysis group, we used 132 present-day Eurasian populations as a donor panel and ran MOSAIC assuming two or three ancestry components (-a 2 or -a 3). Otherwise, we used default parameters. After running MOSAIC, we extracted genomic segments with high posterior probability for being assigned to each ancestry. For each ancestry component, we took a position from each phased haplotype if the posterior

probability from MOSAIC is 0.9 or higher. Then we merged two phased haplotypes from each individual to create ancestry-specific genotypes for each individual and ancestry component. If both alleles passed the threshold, we took the diploid genotype. If only one allele passed the threshold (either the other belongs to other ancestry components or it has low posterior probability), we took the single allele as a pseudo-haploid genotype. This was to maximize per-ancestry genotype coverage.

As an independent validation, local ancestry analysis was inferred with a different method, RFMix v2.03 (Maples et al., 2013) that identifies chunks of contiguous ancestry of a given phased data using a conditional random field (CRF) parametrized by random forest and trained on a reference panel of haplotypes from two or three distinct sets of populations. As for Mosaic, phased 1240KHO modern populations from Europe, Caucasus and the Middle-East were used as Western Eurasian reference (WE) while East Asians and Siberians were used as eastern Eurasian reference (EE). The default 8 generation since admixture was considered as well as default CRF and random forest window sizes. The most likely assignment of subpopulation per CFR region (the ''.msp.tsv'' output) was used for each haploid genome to independently mask the regions assigned to EE or WE. The two haploid masked genomes per each individual were re-collapsed by taking the haploid allele where one region is masked, and taking the diploid call where both regions are assigned to the same ancestry.

We used the ancestry-specific genotype data to run PCA (Figure S2C) and qpWave/qpAdm (Figure S2B). The qpAdm results from both local ancestry methods (MOSAIC and RFMix) were consistent with each other and with the main evidence obtained via the competing 3-way qpAdm models performed on the pseudohaploid data. We otherwise advise caution in interpreting results supported only by these or similar imputation-haplotype-based methods since their application on ancient data is only at the first stages of research and have not been fully validated.

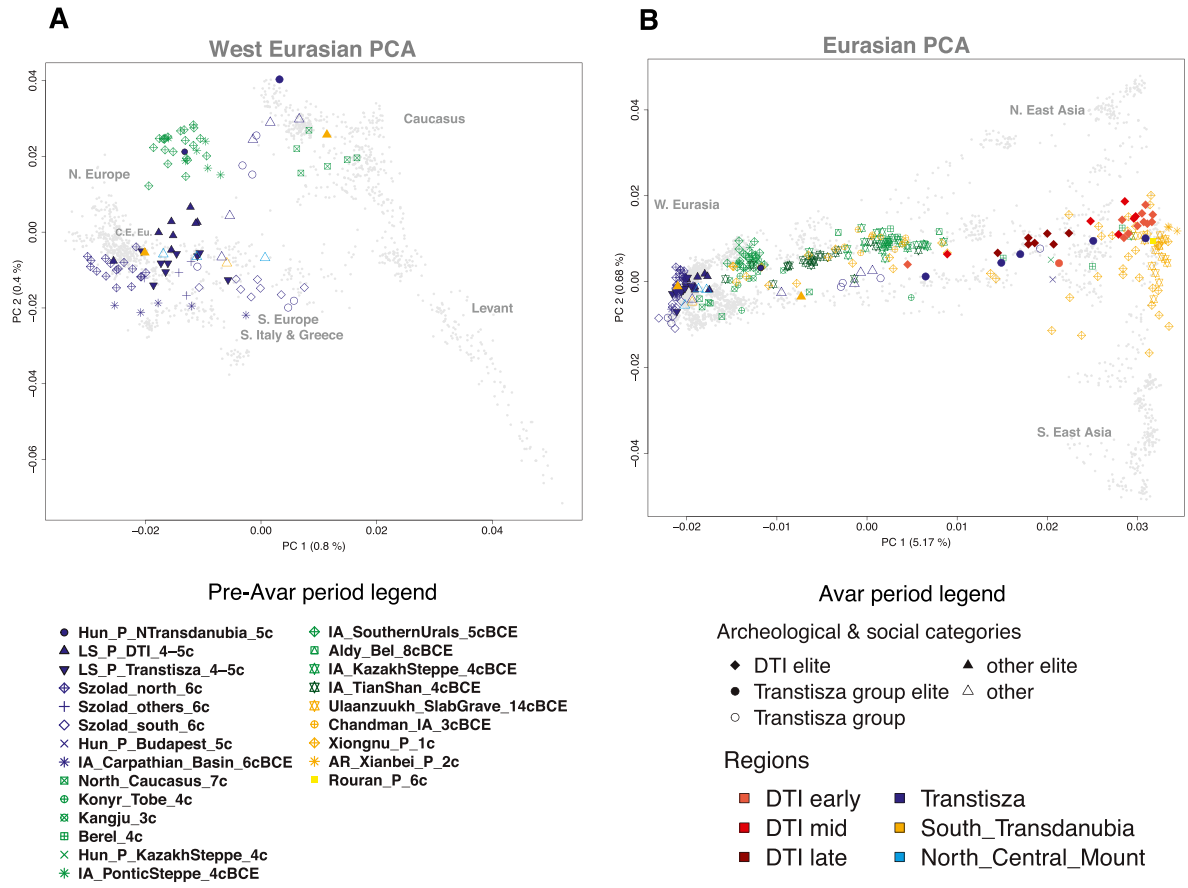
Runs of homozygosity

We tested for the presence of inbreeding in the studied individuals by calculating runs of homozygosity (ROH; i.e. long stretches of homozygous segments along the genome of an individual).

On imputed data we run ROH using the function ''–homozyg'' implemented in PLINK v. 1.9 (Chang et al., 2015) with the following parameters ''–homozyg-window-missing 25'', ''–homozyg-snp 65'' ''–homozyg-kb 2000'' on a dataset priorly filtered for minor allele frequency ''–maf 0.1''. We also applied hapROH (Ringbauer et al., 2021) a new method that allows estimating ROH from pseudo-haploid data (1240K dataset). We studied ROH > 2, 4, 8, 12 and 20 cM. Detected ROH blocks above 4cM were plotted using the python package implemented in hapROH (https://pypi.org/project/hapROH/) (Figure S4A).

CellPress
OPEN ACCESS

Supplemental figures



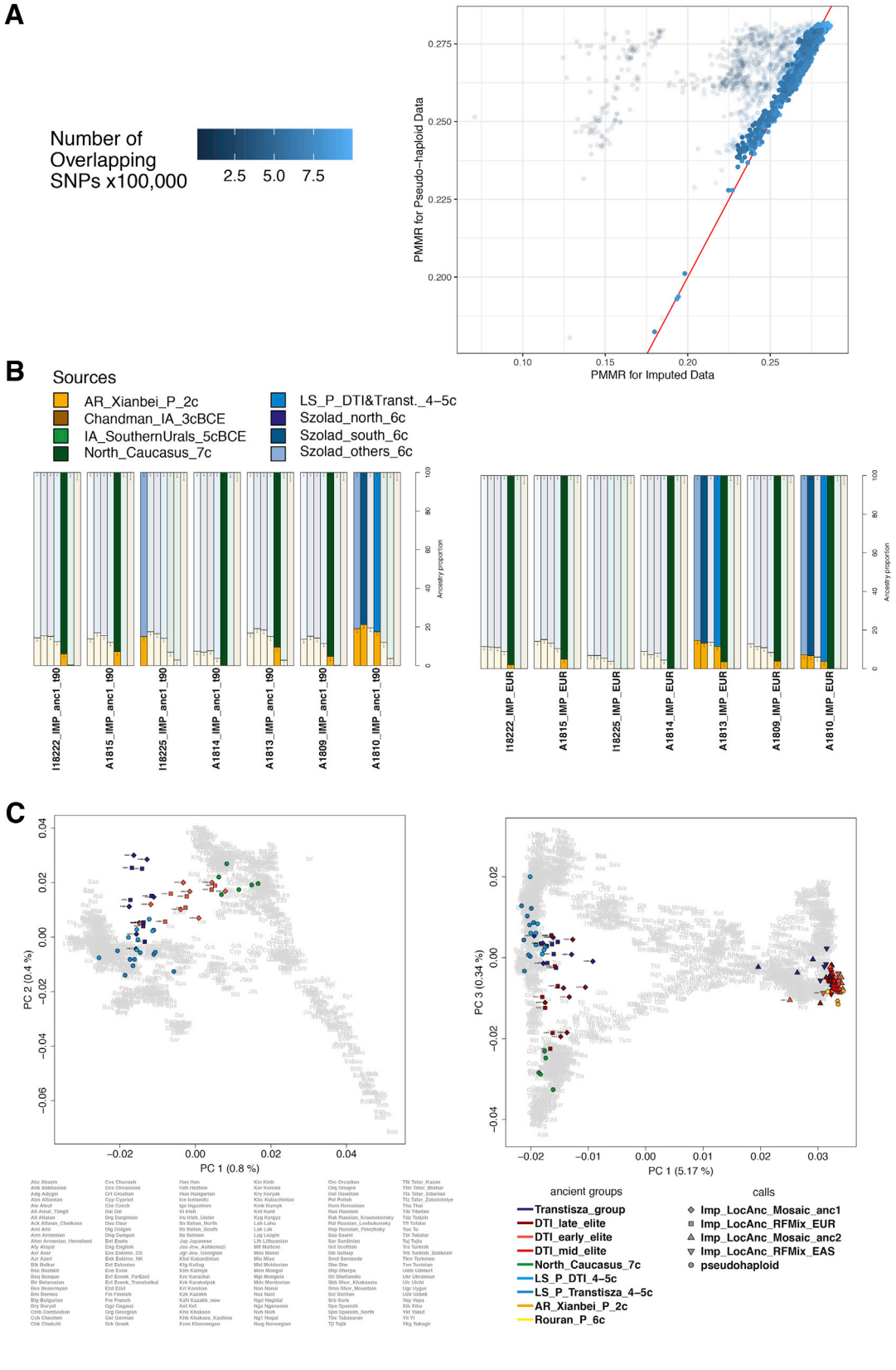
(legend on next page)

Figure S1. West Eurasia and Eurasia PCAs and Most relevant qpAdm models of DTI Avar-period elites/elite associated individuals

(A and B) (A) West Eurasian PCA and (B) Eurasian PCA. The symbols and color scheme are the same as in Figure 2.

(C) From left to right qpAdm models for: DTI early Avar period elite individuals; DTI middle Avar period elite individuals; DTI late Avar period elite individuals. Black boxes highlight the outlier infant (A1817, DTI early Avar period) and child (I18744, DTI middle Avar period).

(D) Three-way competing models of DTI late-Avar-period individuals contrasting local + non-local sources in the same model. A transparency factor is added to the models presenting poor fits ($p < 0.05$), related to Figures 1, 2, and 3 and Tables S1 and S2.



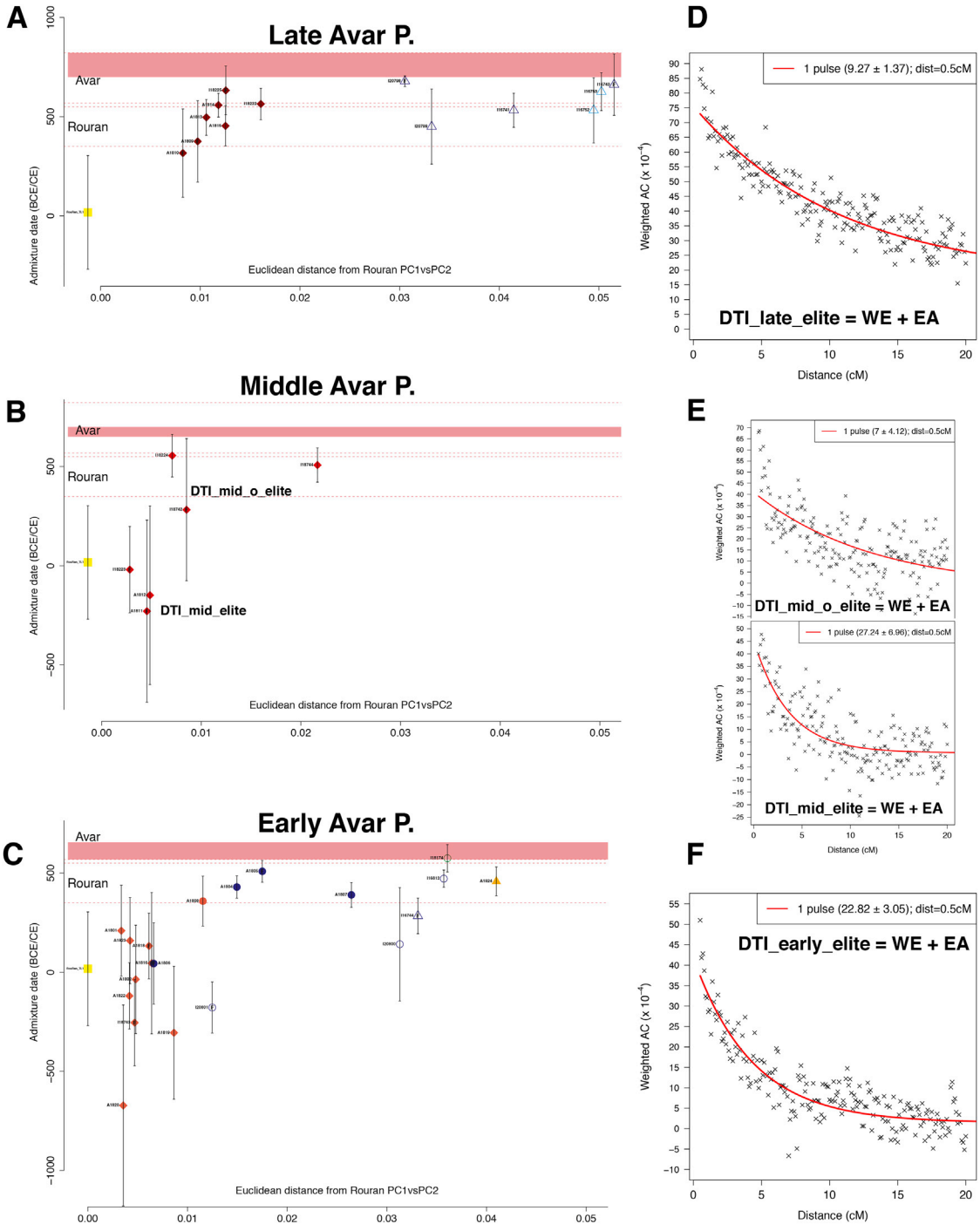
(legend on next page)

Figure S2. Summary of imputation quality control and post-imputation analyses

(A) Scatter plot of pairwise mismatch rate for the pseudohaploid data (x axis) versus the pairwise mismatch rate for the imputed data (y axis). Individuals are filtered to have the proposed coverage cutoff of 1.43×. Points are colored by number of overlapping SNPs in each pairwise mismatch rate calculation. The red line is the y = x line. Transparent points indicate pMMR values for individuals, which were not included because falling below our coverage threshold.

(B) Two-way qpAdm models and PCAs for the West Eurasian component (i.e., masking East Asian ancestry tracts) of DTI late Avar period individuals with Mosaic (left) and RFMix (right).

(C) West Eurasian (left) and East Eurasian (right) PCAs projecting the masked local ancestry tracts of DTI late individuals performed with both methods (Mosaic and RFMix) and pseudohaploid data of individuals representative of local (Sarmatian period) and non-local (North_Caucasus_7C) individuals, related to Figure 3 and Table S1.



(legend on next page)

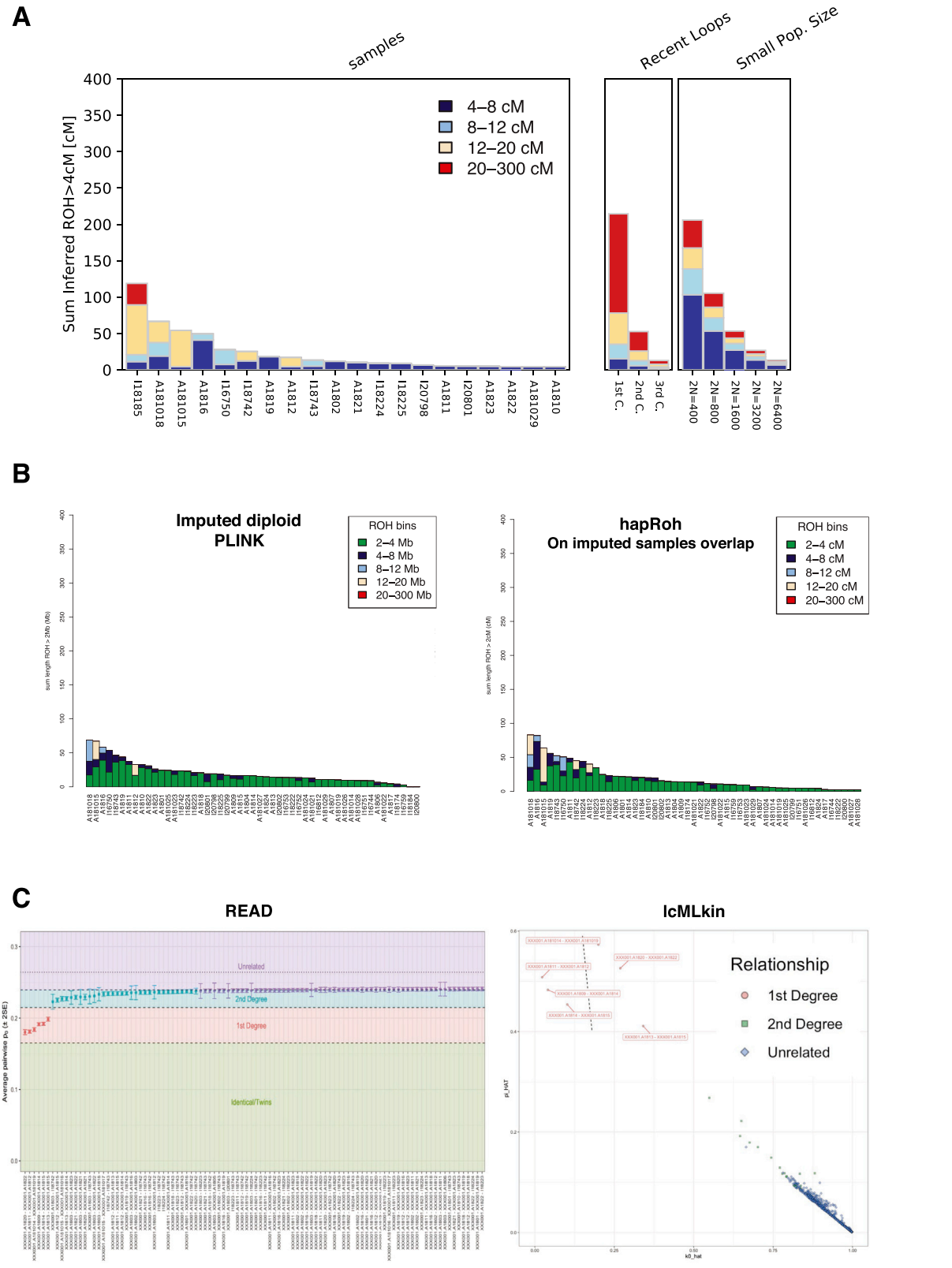
Figure S3. All individual-based Avar-period admixture dates obtained with DATES plotted against Euclidean distance from PC1 and PC2 of the Rouran-period individual

(A) Late Avar period.

(B) Middle Avar period.

(C) Early Avar period. The Rouran-period genome is also shown with its estimated admixture date and at 0 distance from itself on the x axis.

(D–F) Ancestry covariance decay plot for the group-based analyses (WE = western, EA = eastern sources), related to Figure 4.



(legend on next page)

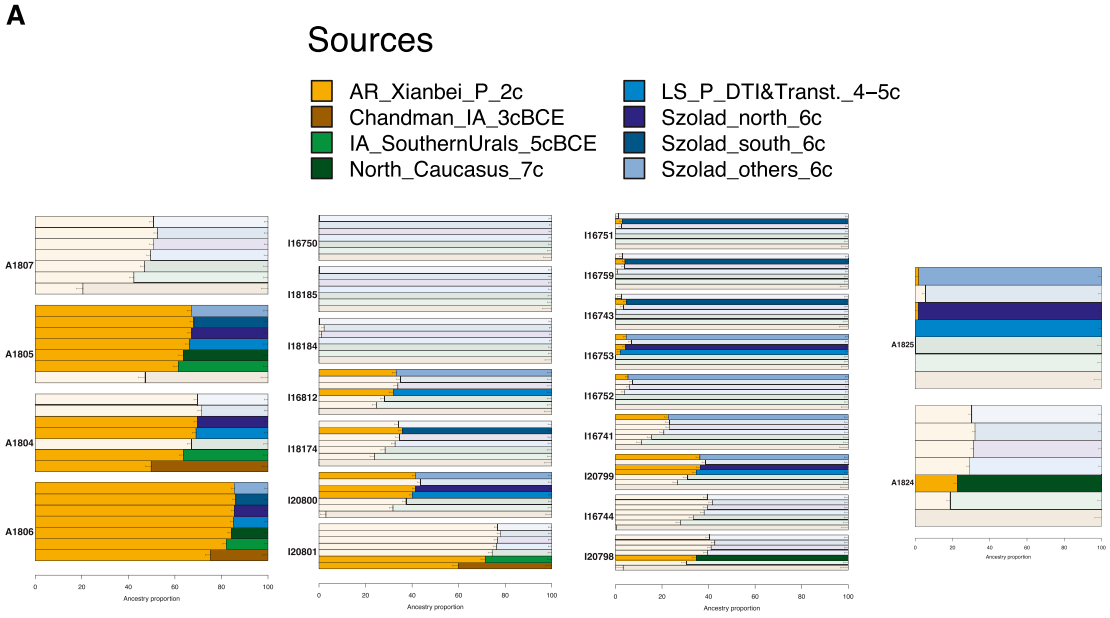
Figure S4. Assessment of genetic inbreeding and relatedness

(A) hapROH analyses reporting only individuals showing runs of homozygosity (ROH) tracts longer than 4 cM grouped in 4 length bins. On the right, the result of simulated ROH patterns corresponding to recent inbreeding or small population sizes as provided by the hapROH pipeline.

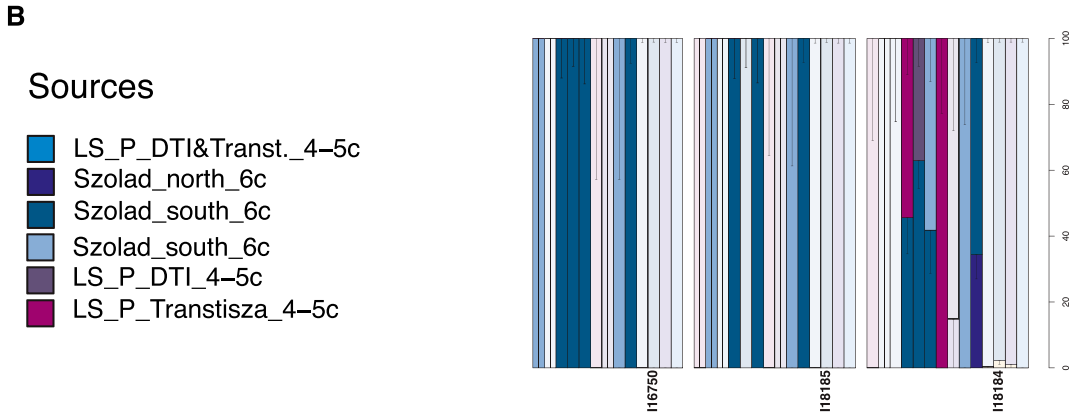
(B) Comparison between ROH on imputed diploid calls run with PLINK "–homozyg" function and hapROH including also the 2–4-cM bins and only overlapping individuals for comparison.

(C) Genetic relatedness estimated with READ (left) and lcMLkin (right), related to Figures 1 and 2.

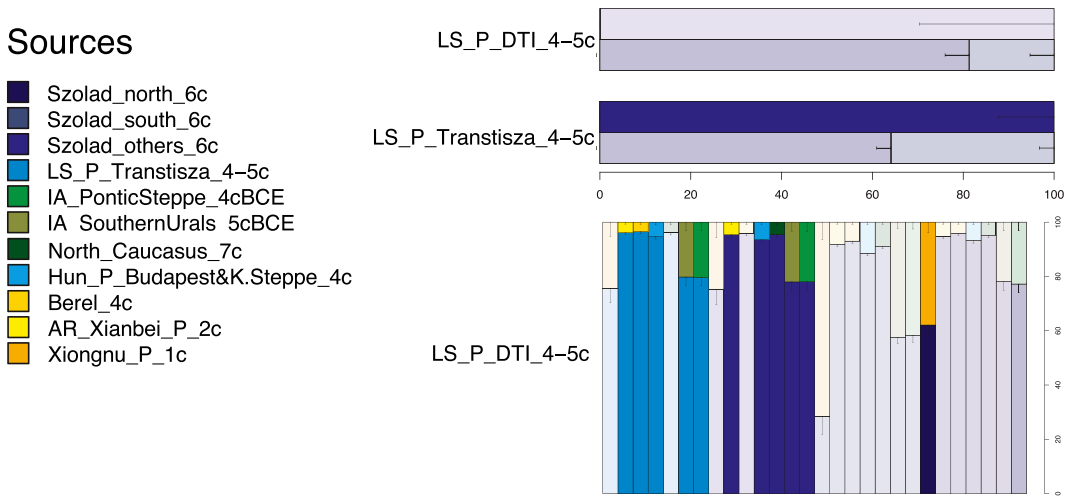
A



B



C



(legend on next page)

Figure S5. qpAdm models for the non-DTI Avar-period individuals and the Sarmatian-period groups

(A) From left to right: two-way models for: early-Avar-period Transtisza group elite associated individuals; early-Avar-period Transtisza group non-elite associated individuals; late-Avar-period non-elite associated individuals; early-Avar-period elite associated (Kölked-Feketekapu site) individuals.

(B) Two-way individual based qpAdm models contrasting different local sources for individuals unresolved with the two-way eastern + western proxies' models. Models for I6750 and I18185 are still non-optimal as they all have infeasible admixture proportions (>>100% for a single source) and large SE despite some having p values > 0.05. Nevertheless, Szolad_south_6c as a unique source shows the less deviant models overall.

(C) qpAdm models for the two Sarmatian period groups: LS_P_DTI_4-5c and LS_P_Transtisza_4-5c. LS_P_Transtisza_4-5c can be modeled without any extra component from the steppe and matches the Szolad_others_6c profile, while LS_P_DTI_4-5c requires additional gene flow from the steppe with different surrogate proxies providing working models, related to Figure 3 and Tables S2 and S4.