Ancient genomes reveal Avar-Hungarian transformations in the 9th-10th centuries CE Carpathian Basin

Dániel Gerber[1]*†, Veronika Csáky[1]†, Bea Szeifert[1]†, Noémi Borbély[1], Kristóf Jakab[1], György Mező[2,3], Zsolt Petkes[4], Frigyes Szücsi[5], Sándor Évinger[6], Csilla Líbor[7], Piroska Rácz[8], Krisztián Kiss[1,9,10,11], Balázs Gusztáv Mende[1], Béla Miklós Szőke[8]‡, Anna Szécsényi-Nagy[1]‡

1) Institute of Archaeogenomics, HUN-REN Research Centre for the Humanities; Budapest, Hungary
2) Konkoly Observatory, HUN-REN Research Centre for Astronomy and Earth Sciences, Budapest, Hungary
3) Wigner Data Center, HUN-REN Wigner Research Centre for Physics
4) Freelancer archaeologist, Budapest, Hungary
5) Szent István Király Museum, Székesfehérvár, Hungary
6) Hungarian Natural History Museum, Budapest, Hungary
7) Hungarian National Museum, Budapest, Hungary
8) Institute of Archaeology, HUN-REN Research Centre for the Humanities; Budapest, Hungary
9) Faculty of Health Sciences, University of Miskolc, Miskolc, Hungary
10) Department of Biological Anthropology, Eötvös Loránd University (ELTE), Budapest, Hungary
11) Department of Anthropology, Hungarian Natural History Museum, Budapest, Hungary

*corresponding author
†these authors contributed to this work equally
‡these authors supervised this work equally

Keywords: Carpathian Basin, Identity by descent, Mediaeval period, population genomics

Abstract

During the Early Medieval period, the Carpathian Basin witnessed significant demographic shifts, notably under Avar dominance for approximately 250 years, followed by the settlement of early Hungarians in the region during the late 9th century CE. This study presents the genetic analysis of 296 ancient samples, including 103 shotgun-sequenced genomes, from present-day Western Hungary. Utilising identity-by-descent (IBD) segment sharing networks, this research provides detailed insights into the population structure and dynamics of the region, spanning the 5th to the 11th centuries CE, with certain microregional foci. Our findings highlight extensive homogenisation and reorganisation processes, discontinuities between Hun, Avar, and Conquest period immigrant groups, alongside the spread and integration of ancestry related to the Hungarian conquerors. This research is among the pioneering efforts to apply IBD analysis to ancient DNA, offering a nuanced understanding of the population dynamics between the 8th and 11th centuries CE in the Carpathian Basin.

Teaser

Genetic formation of mediaeval Hungary, discontinuity between immigrant groups in the first millennium CE

Abbreviations

CB-EUR - genetic cluster/cline in the Carpathian Basin of individuals with European, Caucasian and Near-Eastern ancestry;
CE - Common era;
GHP - Great Hungarian Plain;
IBD - Identity-by-descent;
k_B - degree centrality of IBD connections between archaeological sites;
k_W - degree centrality of IBD connections within archaeological sites;
post-CEE - genetic cluster of individuals with East-Eurasian ancestry chronologically following the Hungarian conquest;
pre-CEE genetic cluster of individuals with East-Eurasian ancestry chronologically preceding the Hungarian conquest;
SNP - Single nucleotide polymorphism;
STR - Short tandem repeat;
UDG - Uracil-dehydrogenase;
ZV - Zalavár.

Introduction

The Migration period in Central Europe had great complexity, which was triggered by a number of Eastern and Western cultural and biological impacts, documented by written sources and previous archaeological studies (Bálint, 2019; Bóna, 2002; Bóna et al., 1993; Pohl, 2018; Szőke, 2021; Vida, 2008; Wolfram, 1987). The Carpathian Basin (including the territory of today's Hungary) has been the target for many archaeogenomic studies (Amorim et al., 2018; Gnecchi-Ruscone et al., 2022; Maróti et al., 2022; Vyas et al., 2023) shedding step-by-step light on population events of this period.

The Carpathian Basin was mostly home to transient groups and polities (*e.g.* Suebs, Goths, Lombards, Gepids, *etc.*) with various ethnic and cultural backgrounds between the fall of the Roman Empire in Pannonia (433 CE) and the arrival of the Avars in 568 CE, when their elite of Northeast-Eurasian origin (Csáky et al., 2020a; Gnecchi-Ruscone et al., 2022; Maróti et al., 2022), along with various East-European groups settled in the region (Pohl, 2018). According to the written and archaeological sources, the population of the Avar Khaganate was culturally enormously diverse (Szőke, 2021, 2014). A 20-years conflict with the rapidly expanding Carolingian Empire disintegrated the Avar Khaganate by 811 CE, which fell into smaller regional units. The former Pannonian provinces, *i.e.* Transdanubia (corresponding to the territory of today's Western Hungary) and the Drava-Sava Interfluve, became a part of the Carolingian Empire, while the Danube-Tisza Interfluve and the Trans-Tisza regions (the latter two hereafter referred to as the Great Hungarian Plain, abbreviated as GHP) remained under the supervision of the Khagan. The Khaganate's former territories had been occupied by the Moravians in the northwest and by the Croatians in the southwest, both under the supervision of the Carolingian emperor. This expansion of the Carolingian Empire was followed by its cultural influence and strategic integration in the corresponding regions during the 9th century CE. This involved the establishment of smaller, isolated counties as administrative units and a deliberate effort to christianise the newly acquired territories (Szőke, 2021).

The power centre of Transdanubia was founded at Zalavár-Vársziget (hereafter Zalavár) by Priwina and Chezil around 840 CE. Its population was presumably supplemented by communities arriving from various regions from the Baltic to the Black Sea. Between 840 and 870, three churches were built here, to where the elites and their servants were being buried. Around 880 the settlement became a royal centre, with a culturally diverse population. Around 896 CE, Prince Braslav was charged by King Arnulf with reinforcing the settlement due to the arrival of the Hungarian conquerors (also referred to as Early Mediaeval Magyars). These conquerors overran Transdanubia in the 10th century CE and settled across the region, presumably displacing only the ruling class in the following century. Zalavár remained a prominent settlement and was recognised as the capital of Somogy and Zala counties, both of which still exist today. The population density increased, and a new Benedictine church was consecrated in 1019 CE, with a cemetery around it populated during the following decades (further information in Supplementary Information section 1).

The Carpathian Basin faced a major cultural transition when the Hungarian conquerors arrived in the region and formed the Hungarian Kingdom by the end of the 10th century CE. This entire century is referred to as the Conquest period. All previous administrative structures were disregarded during the century and new centrums were established (Szőke, 2021). The extent to which the 9th century CE population survived and either coexisted or admixed with the new conquerors remains uncertain. These conquerors along with the local population were ultimately Christianized during the Árpádian era (11th-13th centuries CE).

Research aims and sampling foci

The majority of the ancient DNA samples published from the 6th to 11th centuries CE originate from GHP, whereas Transdanubia almost completely lacks ancient genomes from this period, despite its significant historical events after the 5th century CE. Our primary objective was to achieve more comprehensive coverage of the area's population, with a particular focus on the 8th to 11th centuries CE. This was undertaken to test for population continuity, as well as general structure and connections of the prevailing communities during the transition from the late Avar Khaganate to the beginning of the Árpádian era. We aimed at analysing the diversity and stability of the local population across the Carpathian Basin and the dynamics of admixture from Avar and Hungarian conquest event-related immigrants.

An overall 295 samples were collected from various sites (see Fig. 1.a, Table 1 and Supplementary section S1, Supplementary Table S1), with special attention to Zalavár, Székesfehérvár area and Visegrád (Visegrád-Sibrik hill and Visegrád-Széchenyi street 25 sites), as each represent key aspects of the studied historical transitions. Zalavár site represents the greatest centre (*Mosaburg*) of the Carolingian era, for which we sampled three consecutive archaeological horizons (9th (ZVI), 10th (ZVII) and 11th (ZVIII) centuries CE, respectively). Székesfehérvár itself was a newly formed settlement in the 10th century CE that became the royal centre of the Hungarian Kingdom, where the population context for the preceding centuries are represented through sampling of the neighbouring sites (Supplementary Information section 1). Himod-Káposztás (hereafter Himod) represents a transitional smaller rural settlement, which was sampled for two horizons (9th and 10-11th centuries CE), while Visegrád was a newly formed religious and political centre of the Kingdom in the 11th century CE.

Results

Out of the initially collected 295 human DNA samples, 103 were eligible for deeper shotgun sequencing, which resulted in genomes with ~0.019-2.14× average genomic coverage (mean ~0.86×). The number of detected SNPs from the 1240k panel (Reich, 2021) varies between 18,875-1,008,685 (mean ~563,152), which enables deep analysis of the dataset. In addition, 184 mitochondrial genomes via mitogenome capture and 56 Y-chromosomal STR profiles were created for Zalavár, Himod, Visegrád and Sárbogárd sites (Supplementary Information section 2, Supplementary Table S1). We also resequenced sample Uyelgi21 under the ID

MS20 from the 8th-10th century CE Trans-Uralic Uyelgi site (Csáky et al., 2020b) for its importance in Hungarian prehistory to 1.89× average genomic coverage (952,072 genomic SNP on 1240k, Supplementary Table S1) and used as a proxy to distinguish Northeast-Eurasian and Western-Siberian ancestries and connection systems in our analyses. For further sample information, see Supplementary section S1. In the following sections we co-analysed our dataset only with aDNA samples from the territory of today's Hungary dated between the 5th to the 11th centuries CE (Amorim et al., 2018; Gnecchi-Ruscone et al., 2022; Maróti et al., 2022; Vyas et al., 2023), except for the *f*-statistics, where we involved samples from (Antonio et al., 2024) as well. Accordingly, the complete dataset that was also eligible for IBD (identity-by-descent) analysis consisted of a final set of 416 genomes (Table S5).

Genomic composition in the 5th-11th centuries CE Carpathian Basin

Both the principal component analysis (PCA) plot, projected onto modern West-Eurasian background (Fig. 1.b and Supplementary Figures S3.1.1-3.1.5), as well as the K=6 unsupervised ADMIXTURE analysis (Supplementary Fig. S3.1.6, Supplementary Table S2) shows large diversity among the studied samples from Transdanubia (5th-12th centuries CE), which is also reflected by the uniparental makeup that represents a wide range of Eurasian macrohaplogroups both at mitochondrial and Y-chromosomal level (Supplementary Information section 2).

To describe the variability of the investigated dataset, we first separated individuals bearing only European ancestry through several f_4 analyses (Patterson et al., 2012), and tested these for population structure, further described in Supplementary Information section 3.2. Our analyses revealed three major genetic groups (1-3) and two subgroups of group 3 (3.1 and 3.2) among individuals with only European ancestry that is also reflected by PCA positions (Fig. 2a). After performing several f_4 tests (Supplementary Information section 3.2), we found that these groups represent a cline in the region roughly distributed in a northern-southern axis (Fig. 2, Tables S1 and S5), where groups 1 and 2 are the two extremes and group 3 is generally a mixture or cline between these, hence hereafter we refer to it as European cline 3. This description is a refined version of the Carpathian Basin European diversity characterised by Maróti et al. (2022). We also revealed that Near-Eastern ancestry strongly correlates with group 1 and cline 3.1 (Supplementary Information section 3.2). The genomic makeup of group 1 and cline 3.1, in par with their PCA positions and uniparental makeup, point to strong Late Antique and/or Balkan influence in the region, observed by Olalde et al. too (Olalde et al., 2023), as well as suggesting similar impacts in the territory of today's Hungary compared to the Balkans, during the first half of the first millennium CE. Similarly, group 2 can be paralleled with one or more northern European impacts as it was partially described in Vyas et al. (2023). The f_4 grouping results will be confirmed later by IBD analyses as well (Supplementary Information section 3.2, Fig. S4.3). We refer to the individuals bearing only European, Near-Eastern and Caucasian ancestry CB-EUR hereafter, as the latter two are basically co-occurrent, do not show any specific ingroup separation in regional IBD analyses (see below), and finally, based on Olalde et al. (2023), these appear in the Balkan by the end

of the Roman period, therefore are the part of the autochthonous population makeup by the studied era.

The chronological and geographical distribution (Fig. 2b) of these genetic groups/clines and individuals with non-European ancestry shows differences between Transdanubia and the GHP until the 10th century CE. Roughly half of all investigated samples possessed some non-European ancestry (tested by f_4-statistics), although the prevalence of East-Eurasian ancestry is twice as high in GHP compared to Transdanubia (Fig. S3.2.4, Supplementary Table S4). It is important to note that f_4 is not necessarily eligible to detect trace amounts of ancestries, thus minor East-Eurasian ancestry may be present in the CB-EUR individuals, which, however, did not distort further analyses.

Genomic haplotype segment sharing within the Carpathian Basin

Due to the methodological limitation of the allele frequency based methods (Yüncü et al., 2023), we also applied IBD analysis using *ancIBD* (Ringbauer et al., 2023) on high-quality samples after *GLIMPSE* phasing and imputation (Rubinacci et al., 2021). For details, see Materials and Methods section, and for the results, see Supplementary section 4 and Table S6.

We define degree centralities (corresponding to the number of IBD connections with sum IBD>10 cM, both individually and group-based), which vary across centuries and regions. Applying edge-weighted spring embedded layout on the undirected graph, two major clusters are formed (Fig. S4.1) by individuals with predominant East-Eurasian ancestry, that strongly correlates with chronology (separation of the clusters by the event of the Hungarian Conquest around ~890's CE). The late Kushnarenkovo culture (8th-10th centuries CE) associated individual Uyelgi21 shows an extensive connection system to the post-Conquest individuals with East-Eurasian ancestry (post-CEE) and only one 10 cM connection to the pre-Conquest cluster of individuals with East-Eurasian ancestry (pre-CEE), strengthening the separation of the two. Individuals of CB-EUR show a rather homogeneously scattered network, likely as a result of greater population size (see below). In the following, we redefined these East-Eurasian ancestry clusters according to chronological distribution (pre-CEE: 7th-9th centuries CE, post-CEE: 10th-11th centuries CE). While the chronology and IBD cluster classification of East-Eurasian ancestry individuals mostly correlate, we revealed some exceptions (Table S5, detailed in Discussion).

Most of the post-CEE cluster individuals are situated on a different cline on Eurasia PCAs from the individuals of the pre-CEE cluster, many in the direction of and almost overlapping with Uyelgi21 (Fig. S3.1.8), in line with its extensive IBD connections to them. Regardless of several low segment size connections, we find predominant discontinuity between pre- and post-CEE clusters too (Fig. 3). Since the majority of individuals in the pre- and post-CEE clusters also possess West-Eurasian ancestry, we identified those completely lacking this component in Supplementary Information Section 3.2. This was done to determine whether the sporadic connections between pre- and post-CEE clusters are through CB-EUR

individuals or common East-Eurasian ancestry. Accordingly, we observed the latter through a few connections but these are accumulating among post-CEE individuals from the second half of the 10th century CE in a few scattered areas, pointing to sporadic and regionally different survival of these lineages into the post-CEE cluster, and ultimately local admixture and not common eastern ancestry.

The IBD connections of the pre-CEE cluster to CB-EUR are way more sparse compared to those of the post-CEE cluster (the average individual connection to CB-EUR for pre-CEE is 0.54 and for post-CEE is 1.39, see Fig. 3). Contrary, the average within cluster degree centralities for pre-CEE (7) and post-CEE (6.7) are quite similar. The kinship level (1st-5th degree) connection network is remarkably different between the Avar and Conquest periods. Such connections of the pre-CEE cluster remain within except one sample (KDA-188, Kiskundorozsma site) and are all restricted to the same site. Contrary, while the post-CEE cluster connections are mostly restricted to the same site too, primarily occur between CB-EUR and post-CEE individuals. Another striking feature is the relative isolation of the pre-CEE cluster: despite extensive admixture with West-Eurasians (based on PC1 positions), the paucity of CB-EUR connections suggests a non-local source for this component, in line with the observations of (Gnecchi-Ruscone et al., 2024).

The IBD results are more conspicuous in the light of the effective population size (N_e) of the pre-CEE cluster compared to the post-CEE cluster, calculated with hapROH (see Methods and Table S8). We compared pre- and post-CEE clusters' N_e to their contemporaneous subset of CB-EUR individuals to get a proportion of their effective size rather than estimating actual population sizes due to the lack of solid historical or anthropological demographic data regarding the period (Juhász, 2015). Accordingly, ~18.5% of the population of the Avar period belonged to the pre-CEE cluster, while only ~10.5% of the population of the 10-11th centuries CE belonged to the post-CEE cluster. Additionally, the N_e estimates show a small but definite increase towards the end of the millennium for the CB-EUR (Table S8).

The IBD connections of CB-EUR (Fig. S4.3) reflect the f_4-based classification but also reveal that the different sections of the CB-EUR cline probably consist of different populations, *e.g.* those belonging to European group 2 show discontinuity pre- and post-7th century CE (Fig. S4.4) and thus represent different groups that only share similar regional genetic makeup.

Population genetic events in the 5th-11th centuries CE

We explored weighted degree centrality (*i.e.* the number of connections denoted by k between (k_B) and within (k_W) sites, weighted by number of samples per site) across the 5th-12th centuries CE (Fig. 4, Methods). We calculated k_B and k_W values separately within Transdanubian and GHP macroregions, and also combined the analyses for both areas (denoted as Merged in Fig. 4) to reveal both micro- and macroregional patterns.

The cumulative k_B values (*i.e.* where intersite connections from the preceding centuries were added) remained consistent only for the CB-EUR cluster (Fig. 4a) until the turn of the

9th-10th century CE. Then, most likely due to the Hungarian conquest, Transdanubian sites show a high increase in intersite connections contrasting the GHP. The merged values increasing the Transdanubian ones, on the other hand, suggests movement from GHP to Transdanubia. The same process can be observed when we add all samples, *i.e.* those with East-Eurasian ancestry (Fig. 4b) to this analysis, but the pre-Conquest period shows a completely different makeup for GHP. In this case, a high number of intersite connections can be observed, in line with the observations of Gnecchi et al. (2024), who suggested sedentary lifestyle along with intensive female exogamy within the eastern origin strata of the Avar period from the 7th century CE onwards. These results are also pointing out the similar social makeup between Transdanubian and GHP local (CB-EUR) populations, and their highly different social organisations compared to the communities included in the pre-CEE cluster. The lower "merged" values compared to GHP also suggests that GHP was rather isolated from Transdanubia before the 10th century CE.

The non-cumulative approach of the same analysis (*i.e.* only investigating between sites (k_B) connections in the corresponding period, Fig. 4c) revealed two major peaks, signalling two waves of intensive movement and admixture between sites. The first wave can be correlated with the changing Avar society in the 7th century, while the second wave is consistent with the reorganisations fueled by the Hungarian conquest in the turn of the 9th-10th centuries CE. However, this observation contradicts those of Gnecchi et al. (2024) on continuous female exogamy after the 7th century CE. Also, some 7th century CE sites, such as Kunpeszér, Kunszállás from Maróti et al. (2022) and Gnecchi et al. (2022) and even the few samples from a large burial site of Madaras (Fig. 3), show high degree of within site connections, regardless of biological sex.

The most interesting picture, however, can be observed in Fig. 4d, which shows the intra-community (k_W) changes throughout the centuries. The steadily increasing connections within the sites (k_W) from the first Avar settlements until the 9th century CE suggest a degrading biological network between communities in this period. In the Conquest period (10th century CE), however, k_W greatly dropped simultaneously with the increase of k_B, signalling the previously observed intensive movement between sites. Interestingly, in Transdanubia, the decrease in k_W values shows a few generations delay compared to the GHP, correlating with archaeological-historical data that the Hungarian conquest was a sequence of events in the Carpathian Basin from eastern to western territories (Langó, 2017). After the Hungarian conquest, in both regions k_W values tend to slowly increase in the following centuries, pointing to a more settled lifestyle of prevailing communities.

Discussion

Fluctuations within the CB-EUR cluster throughout the 5th-10th centuries CE

The chronologically changing makeup of distant segments of the CB-EUR genetic cline defined by our analyses reveals complex population processes in the Carpathian Basin during

the second half of the first millennium CE. European group 2 individuals are rather abundant in Transdanubia, along with related cline 3.2. Common elements between regions are the proportional decline of European genetic group 1 and Near-Eastern ancestry, as well as the proportional enrichment of (i.e. the integration of groups 1 and 2 into) cline 3 towards the 11th century CE. The proportional similarities of the two regions in the 10th century CE suggests an amalgamation process of different groups across the studied area, which we confirmed by IBD segment connection dynamics as well.

We highlight that the presence of clinal similarity between different regions or horizons does not necessarily indicate group identity, even at a biological level, as European group 2 shows discontinuity pre- and post-7th century CE based on IBD results (Fig. S4.4). This observation underscores that PCA and evolutionary-oriented techniques, such as ADMIXTURE and *f*-statistics based methods, are limited to correctly characterise populations in the Mediaeval complexity, and thus conclusions drawn from such results must be approached with great caution. However, certain trends should not be overlooked, *e.g.* in the pre-Conquest period the presence of a remnant Romanised population around Csákberény based on archaeological findings (Supplementary Information section 1.8) is well paralleled with the group1/Near Eastern genetic makeup of the individuals from this site (Fig. 3, Supplementary Information section 3, Results). The accumulation of certain elements further suggests some local isolation in Himod I with group 2 elements and GHP sites Kiskundorozsma and Jánoshida (Maróti et al., 2022) with group 1 elements, although these are less paralleled with archaeological background (Fig. 3) and their genetic isolation is shaded by connection to a pre-CEE individual in Kiskundorozsma.

Dynamics of the communities with East-Eurasian ancestry before and after the Hungarian Conquest period

Recent researches by Gnecchi-Ruscone et al. (2022, 2024) and Maróti et al. (2022) has illuminated the East-Eurasian ancestry being predominant among the early Avar Khaganate's elite, indicating that these groups were a significant genetic source of immigrants (Fig. S3.1.8). Therefore we conclude that the pre-CEE cluster defined in this study mostly represent the East-Eurasian core and genetically related individuals of the Avars, although we acknowledge that ethnic identities and genetic categories used do not correspond directly. Similarly, the post-CEE cluster can strongly be associated with the Hungarian conquerors and their descendants. The remarkable discontinuity between pre- and post-CEE clusters is shaded by the sporadic IBD connections, which almost exclusively accumulate among the second half of the 10th century CE cemeteries, pointing to the limited local survival of pre-CEE cluster lineages. The only exception is Karos cemetery, which represents one of the earliest Hungarian conqueror cemeteries in the Carpathian Basin but also shows the strongest IBD connections to the pre-CEE cluster. On the other hand, the diverse genetic backgrounds observable here indicate the varied origins of this community's members, framed within a military unit as evidenced by the ratio of armed male burials and the diverse cultural elements found in the cemetery material (Révész, 1996). Also, contrary to the suggested continuity of 5th century CE Hun period East-Eurasian genetic ancestry into the Conqueror cluster by

Maróti et al., 2022, our data indicate an absence or loss of this ancestry among post-CEE individuals in the studied region (Fig. S4.5).

The IBD connections of the Hungarian conquerors to CB-EUR demonstrates an immediate mingle with the locals, while the rarity of such connections between the eastern Avars and the parallel CB-EUR (Fig. 3) is striking, especially considering these groups relative population size. These results, however, can have many interpretations. The genetic isolation of the Avar elite that has been shown in previous studies (Csáky et al., 2020a; Gnecchi-Ruscone et al., 2024, 2022) are now verified by our IBD analyses as well, and highlight the different social organisations of the communities belonging to the pre- and post-CEE clusters. However, we cannot confirm the observations of Gnecchi et al. (2024) on the extended exogamy as a general residence pattern of the Avar society throughout the centuries. Based on the results summarised in Fig. 4, the Avar communities show a wave of between-site connections (likely driven by female exogamy) only in the 7th century CE, which they did not maintain towards the 9th century CE. This observation is supplemented by the within-site kinship-level connections of late Avar period sites, such as Árkus and Szarvas (Fig. 3), regardless of biological sex. It is notable that at the investigated sites by Gnecchi et al. (2024), many females did not connect to the studied pedigrees and many pedigree forming females are missing from the family trees. In a general multi-site (macroregional) model that we applied, the overall picture shows a different social makeup, although we can not exclude exogamy from closely neighbouring villages. This idea is also supported by the strontium isotope results from Gnecchi et al. (2024), who did not describe large-scale movement between distant villages. The relative isolation (Fig. 3) found among the Avars offers an explanation for their disappearance and the succession of the Hungarians both culturally and genetically. Additionally, these results also point to the West-Eurasian (likely Eastern-European, steppe and/or Caucasian) admixture in the pre-CEE cluster originating mostly outside the Carpathian Basin, in line with the observations of Gnecchi et al. (2024). This points to a prevailing West-Eurasian genetic substrate during the Avar period that can currently only be traced through admixture into the elite East-Eurasian families.

Microregional genetic trends in Transdanubia

Focusing on the IBD network of Transdanubia, we grouped sites according to their ecological and chronological distribution (Fig. 5, Supplementary Information section 4). Despite the relatively small number of samples compared to the presumed large populations in the past, the Zalavár sample set demonstrates continuity throughout the 9th to 11th centuries CE. Two individuals from the early 7th century CE occupation of Szólád (AV1 and AV2, first degree relatives) are superconnected to individuals from most of the studied sites, whereas the Zalavár connections are the most apparent. They were recovered from deposit fillings of a Langobard period grave without any artefacts (Amorim et al., 2018). Despite minimal archaeological information they show the presence of a genetically successful (and likely large) substrate in the region that probably arrived around the 7th century CE. Given this genetic substrate's presence centuries before *Mosaburg* was founded, we can not confirm the written sources that state the entire settlement's population originated outside the Carpathian

Basin, which points to exaggerations in historical records. Nevertheless, the mtDNA haplogroup C1e (AHP20) that was so far only recovered from Icelandic populations (Ebenesersdóttir et al., 2011), or the Y-chromosomal haplogroups I1-Y6357 (AHP16) and R1a-FGC11904 (AHP18), which are primarily Northwest-European origin lineages according to Yfull.org, may indicate some foreign connections in the Caroling horizon of Zalavár.

The 9th century CE phase of the northwestern Transdanubian Himod site also shows connections to AV1-2 (7th century CE, Szólád site, Transdanubia) and even the accumulation of CB-EUR group 2 elements as it was highlighted previously. Connections between the pre- and post-Conquest phases (I-II) of the Himod site are sporadic and differences in admixture proportions (Fig. S3.1.7) and uniparental makeup suggest only limited genetic continuity between the horizons. Instead, these results rather reflect Hungarian conqueror occupation (with incoming new genetic components connected to GHP and the post-CEE cluster too) and a subsequent admixture process in Himod II. The northeastern Transdanubian Visegrád site shows more IBD connections to contemporary groups, rather than preceding ones in other regions, also confirming the written sources that describe it as a newly formed community (Kovács, 2022), although future comparative IBD analyses with other European areas are required to fully understand the composition of this society. In the thoroughly analysed Székesfehérvár area, we see a different picture from that of the Zalavár. A Late-Antiquity-type population persisted in this region, as we described in the previous sections, which could even be found around Sárbogárd in the 10th century CE, although the appearance of Eastern elements after the Hungarian conquest is way more prominent in the area. Despite the larger number of samples from southeastern Transdanubia, signs of local succession are less pronounced in this area, consistent with expectations of community restructuring after the Hungarian conquest, described by the IBD degree centrality changes throughout the centuries. Overall, the microregional evaluation reveals spatially different histories even between communities of close geographical proximity, highlighting the importance of the prudent fusion archaeological and biological perspectives when interpreting such datasets.

Changing communities in the late 9th - early 10th century Carpathian Basin

The observed changes in connections between and within sites throughout the 5th-11th centuries reflect both Avar and Hungarian conqueror occupation and the amalgamation process of CB-EUR groups 1 and 2 into cline 3 towards the end of the millennium. Our analyses revealed that the GHP was genetically rather isolated from Transdanubia before the 10th century CE, but the Hungarian conquest fueled intensive population movements and admixtures within the Carpathian Basin, on par with historical records (Szőke, 2021). This process is also reflected by the occurrence of relatives in geographically distant cemeteries after the Hungarian conquest, *e.g.* individuals KOZ018 (Himod, Transdanubia, 10-11th century CE) and TCS-18 (Tiszanána, GHP, 10th century CE) are a father-son pair, KOZ026 (Himod, Transdanubia, 10-11th century CE) and SH-81 (Sárrétudvari, GHP, 10th century CE)

are 3rd-4th degree relatives (Table S6). We observed a slight delay in this process in Transdanubia compared to GHP, another difference between these two microregions that highlight the importance of considering ecological barriers (in this case the river Danube) in reconstructing past population events. We also found evidence for pre-conquest appearance of individuals with Uralic ancestry. In phase ZVI (dated to 850-890), the adult male AHP21 is strongly connected to multiple individuals in the post-CEE cluster. This individual's genomic composition is highly similar to Uyelgi21 (and also has strong IBD connections to it; see Table S6 and Fig. S3.1.8), and is the 5th-degree relative of adult male KeF1-10936 from Kenézlő-Fazekaszug (GHP), whose dating cannot be defined more precisely than 900-960 CE.

Additionally, AHP29 from the Carolingian era (ZVI) shares IBD segments with the Conqueror cluster too, although its connections are much weaker compared to AHP21's. Both burials firmly predate the Hungarian conquest by archaeological stratigraphy (Supplementary Information section 1). These signal sporadic appearances of Hungarians during the 860-870 CE period in Transdanubia, who likely served as mercenaries at the court of Priwina (based on historical records, see Supplementary Information section 1.1).

Limitations and final remarks of this study

Our results revealed complex population histories and both macro- and microregional events in the Carpathian Basin. We believe that the combination of our and the database data, as well as the methods we used provide enough statistical strength to the statements of this study. Since we did not include comparative dataset from other countries, as the main goal of our study was the local characterisation of population events, we are aware that *e.g.* ancestry shift within CB-EUR group 2, the origins of the Late-Antiquity-type population or the better characterisation of the post-CEE cluster left open questions that can be explained by exhaustive comparison to data of other populations in the future. We believe that further sampling from East-Central Europe is required to reveal such delicate questions, and we aim to do so for subsequent studies.

Materials and Methods

Sampling

Ancient DNA laboratory work

Samples are processed in a dedicated laboratory aDNA facility at the Institute of Archaeogenomics, Research Centre for the Humanities, HUN-REN. Sample preparation has taken place according to (Szeifert et al., 2022). Bone powders were produced based on (Lipson et al., 2017; Sirak et al., 2017), DNA extractions were performed following the methods described in (Dabney et al., 2013; Rohland et al., 2018) and the DNA libraries were created using half-UDG treatment (Rohland et al., 2015). Both DNA extraction and library preparation were performed manually or automatically with Biomek i5 Liquid Handling

Workstations. Initial screenings were performed on an inhouse Illumina MiSeq platform, eligible samples were then processed on Illumina platforms (Applied Biotechnologies and Novogene) for deep sequencing, for all results, see Supplementary Table S1.

We performed mtDNA hybridisation capture (Rohland et al., 2022; Szeifert et al., 2022) and Y chromosome STR analyses using AmpFlSTR® Yfiler® PCR Amplification Kit (Applied Biosystems). Y chromosome haplotype analysis was performed with GeneMapper® ID Software v3.2.1 (Applied Biosystems).

Bioinformatic analyses

The *PAPline* package with default settings (Gerber et al., 2023) were used for Illumina paired-end sequencing reads mapping to *Homo sapiens* reference genome GRCH37.p13, pre- and post-filtering steps and mitochondrial sequences calls, for results, see Supplementary Table S1. Y chromosome haplogroups were classified using the *Yleaf* v2.2 software (Ralf et al., 2018).

The ratio of male X-chromosomal contamination was measured, with both *ANGSD* (version 0.939-10-g21ed01c, htslib 1.14-9-ge769401) (Korneliussen et al., 2014) and hapCon (hapROH package version 0.60) (Huang and Ringbauer, 2022). The starting file for the former method was a *samtools mpileup* (version 1.10, htslib 1.10.2-3ubuntu0.1) output with mapping quality 25 and base quality 30. All the parameters used were the default ones recommended in the official documentation. Similarly, the chrX_1240k.bed and chrX_1240k.hdf5 files used were supplied by the developers with the software package.

ANGSD doCounts was run on the X-chromosomal region from position 5500000 to 154000000, with mapping and base qualities 25 and 30 respectively. The resulting file was processed then with the toolkit's contamination executable, with the following parameters: -b 5500000 -c 154000000 -d 3 -e 100 -p 1. The required HapMapChrX.gz file (-h parameter) was the one shipped with the software.

Kinship relations were estimated using *READ* (Monroy Kuhn et al., 2018) and *KIN* (Popli et al., 2023), and where available, supplemented by IBD connections, for results, see Supplementary Table S2). For PCA, Admixture and *f*-statistics the dataset was merged with the AADR database v54 (Mallick and Reich, 2023). PCA was calculated using the *smartpca* software (Patterson et al., 2017) with the 590k SNP panel (Mallick and Reich, 2023), for admixture analysis we used the *ADMIXTURE* software (Alexander et al., 2009) with the 1240k SNP panel (Mallick and Reich, 2023). For *f*-statistics, we used the *Admixtools* software package (Maier et al., 2021; Patterson et al., 2012). For the *ADMIXTURE* analysis, we used a dataset which contains genomic data of archaic age individuals from European, Near Eastern, Central Asian, and Mongolian populations. We filtered out from the closely related individuals and those samples that were not treated with UDG.

We used hapROH (Ringbauer et al., 2021) to detect runs of homozygosity (ROH) that indicate consanguinity of parents and small population sizes. The *Ne* vignette of hapROH was used to estimate effective population sizes (N_e) with confidence intervals for f_4 statistics-based genetic groups considering also chronological phases. 95% Confidence Intervals were fitted via both the likelihood profile (calculating the log likelihood for a large number of $2N_e$, and looking for the interval 1.92 log-likelihood units down from the Maximum Likelihood) and using the curvature of the likelihood (the so called Fisher Information matrix).

Individuals with >250k SNPs on the 1240k panel were targeted for IBD analysis. Imputation was performed using *GLIMPSE* software (Rubinacci et al., 2021), following the recommendations of the authors and using the v1.1.1 versions of the precompiled static binaries. The reference panel for the imputation (1000 Genomes Project (30X from NYGC) on hg19/GRcH37) was provided by Ali Akbari (Harvard Medical School). Genotype likelihoods of the samples were computed with a standard method involving bcftools *mpileup* and *bcftools call* v1.10.2 (Li et al., 2009). Base- and mapping quality (flags *-Q* and *-q*) were both set to 20. As a difference from the official workflow, chromosomal level imputation regions were determined on a per sample basis, with the GLIMPSE_chunk_static executable, using default parameters (*--window-size 2000000 --buffer-size 200000*). To identify the IBD segments, we used *ancIBD* (version 0.5) (Ringbauer et al., 2023) with the same process and parameters described in the software's official documentation. The fraction of 1240k variants above 0.99 genotype probability on Chromosome 3 must have been at least 0.8 after imputation to include a sample for IBD calling. All results were visualised using *Cytoscape* (Shannon et al., 2003). We restricted connections down to 1x10 cM, due to the high number of false-positive hits at short-segments resulting from technical difficulties (*e.g.* phasing errors) and the sparse and asymmetric background database for imputation.

We also measured degree centrality (*k*, using connections between 10 and 1400 cM to exclude close relatives, as we were interested in the population-level processes), divided by the sample number per batch to obtain a relative metric.

References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. https://doi.org/10.1101/gr.094052.109

Amorim, C.E.G., Vai, S., Posth, C., Modi, A., Koncz, I., Hakenbeck, S., La Rocca, M.C., Mende, B., Bobo, D., Pohl, W., Baricco, L.P., Bedini, E., Francalacci, P., Giostra, C., Vida, T., Winger, D., von Freeden, U., Ghirotto, S., Lari, M., Barbujani, G., Krause, J., Caramelli, D., Geary, P.J., Veeramah, K.R., 2018. Understanding 6th-century barbarian social organization and migration through paleogenomics. Nat. Commun. 9, 3547. https://doi.org/10.1038/s41467-018-06024-4

Antonio, M.L., Weiß, C.L., Gao, Z., Sawyer, S., Oberreiter, V., Moots, H.M., Spence, J.P., Cheronet, O., Zagorc, B., Praxmarer, E., Özdoğan, K.T., Demetz, L., Gelabert, P., Fernandes, D., Lucci, M., Alihodžić, T., Amrani, S., Avetisyan, P., Baillif-Ducros, C.,

Bedić, Ž., Bertrand, A., Bilić, M., Bondioli, L., Borówka, P., Botte, E., Burmaz, J., Bužanić, D., Candilio, F., Cvetko, M., De Angelis, D., Drnić, I., Elschek, K., Fantar, M., Gaspari, A., Gasperetti, G., Genchi, F., Golubović, S., Hukeľová, Z., Jankauskas, R., Vučković, K.J., Jeremić, G., Kaić, I., Kazek, K., Khachatryan, H., Khudaverdyan, A., Kirchengast, S., Korać, M., Kozlowski, V., Krošláková, M., Kušan Špalj, D., La Pastina, F., Laguardia, M., Legrand, S., Leleković, T., Leskovar, T., Lorkiewicz, W., Los, D., Silva, A.M., Masaryk, R., Matijević, V., Cherifi, Y.M.S., Meyer, N., Mikić, I., Miladinović-Radmilović, N., Milošević Zakić, B., Nacouzi, L., Natuniewicz-Sekuła, M., Nava, A., Neugebauer-Maresch, C., Nováček, J., Osterholtz, A., Paige, J., Paraman, L., Pieri, D., Pieta, K., Pop-Lazić, S., Ruttkay, M., Sanader, M., Sołtysiak, A., Sperduti, A., Stankovic Pesterac, T., Teschler-Nicola, M., Teul, I., Tončinić, D., Trapp, J., Vulović, D., Waliszewski, T., Walter, D., Živanović, M., Filah, M.E.M., Čaušević-Bully, M., Šlaus, M., Borić, D., Novak, M., Coppa, A., Pinhasi, R., Pritchard, J.K., 2024. Stable population structure in Europe since the Iron Age, despite high mobility. eLife 13, e79714. https://doi.org/10.7554/eLife.79714

Bálint, C., 2019. The Avars, Byzantium and Italy. A study in chorology and cultural history, Varia archaeologica Hungarica. Archaeolingua, Budapest, Hungary.

Bóna, I., 2002. Les Huns: le grand empire barbare d'Europe, IVe–Ve siècles (tr. Katalin Escher). Paris, France.

Bóna I., Cseh J., Nagy M., Tomka P., Tóth Á., 1993. Hunok - Gepidák - Langobardok. Magyar Őstörténeti Könyvtár, Szeged, Hungary.

Csáky, V., Gerber, D., Koncz, I., Csiky, G., Mende, B.G., Szeifert, B., Egyed, B., Pamjav, H., Marcsik, A., Molnár, E., Pálfi, G., Gulyás, A., Kovacsóczy, B., Lezsák, G.M., Lőrinczy, G., Szécsényi-Nagy, A., Vida, T., 2020a. Genetic insights into the social organisation of the Avar period elite in the 7th century AD Carpathian Basin. Sci. Rep. 10, 948. https://doi.org/10.1038/s41598-019-57378-8

Csáky, V., Gerber, D., Szeifert, B., Egyed, B., Stégmár, B., Botalov, S.G., Grudochko, I.V., Matveeva, N.P., Zelenkov, A.S., Sleptsova, A.V., Goldina, R.D., Danich, A.V., Mende, B.G., Türk, A., Szécsényi-Nagy, A., 2020b. Early medieval genetic data from Ural region evaluated in the light of archaeological evidence of ancient Hungarians. Sci. Rep. 10, 19137. https://doi.org/10.1038/s41598-020-75910-z

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. 110, 15758–15763. https://doi.org/10.1073/pnas.1314445110

Ebenesersdóttir, S.S., Sigurðsson, Á., Sánchez-Quinto, F., Lalueza-Fox, C., Stefánsson, K., Helgason, A., 2011. A new subclade of mtDNA haplogroup C1 found in icelanders: Evidence of pre-columbian contact? Am. J. Phys. Anthropol. 144, 92–99. https://doi.org/10.1002/ajpa.21419

Gerber, D., Szeifert, B., Székely, O., Egyed, B., Gyuris, B., Giblin, J.I., Horváth, A., Köhler, K., Kulcsár, G., Kustár, Á., Major, I., Molnár, M., Palcsu, L., Szeverényi, V., Fábián, S., Mende, B., Bondár, M., Ari, E., Kiss, V., Szécsényi-Nagy, 2023. Interdisciplinary Analyses of Bronze Age Communities from Western Hungary Reveal Complex Population Histories. Mol. Biol. Evol. 40. https://doi.org/10.1093/molbev/msad182

Gnecchi-Ruscone, G.A., Rácz, Z., Samu, L., Szeniczey, T., Faragó, N., Knipper, C., Friedrich, R., Zlámalová, D., Traverso, L., Liccardo, S., Wabnitz, S., Popli, D., Wang, K., Radzeviciute, R., Gulyás, B., Koncz, I., Balogh, C., Lezsák, G.M., Mácsai, V., Bunbury, M.M.E., Spekker, O., Le Roux, P., Szécsényi-Nagy, A., Mende, B.G., Colleran, H., Hajdu, T., Geary, P., Pohl, W., Vida, T., Krause, J., Hofmanová, Z., 2024. Network of large pedigrees reveals social practices of Avar communities. Nature. https://doi.org/10.1038/s41586-024-07312-4

Gnecchi-Ruscone, G.A., Szécsényi-Nagy, A., Koncz, I., Csiky, G., Rácz, Z., Rohrlach, A.B., Brandt, G., Rohland, N., Csáky, V., Cheronet, O., Szeifert, B., Rácz, T.Á., Benedek, A., Bernert, Z., Berta, N., Czifra, S., Dani, J., Farkas, Z., Hága, T., Hajdu, T.,

Jászberényi, M., Kisjuhász, V., Kolozsi, B., Major, P., Marcsik, A., Kovacsóczy, B.Ny., Balogh, C., Lezsák, G.M., Ódor, J.G., Szelekovszky, M., Szeniczey, T., Tárnoki, J., Tóth, Z., Tutkovics, E.K., Mende, B.G., Geary, P., Pohl, W., Vida, T., Pinhasi, R., Reich, D., Hofmanová, Z., Jeong, C., Krause, J., 2022. Ancient genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites. Cell 185, 1402-1413.e21. https://doi.org/10.1016/j.cell.2022.03.007

Héder, M., Rigó, E., Medgyesi, D., Lovas, R., Tenczer, S., Török, F., Farkas, A., Emődi, M., Kadlecsik, J., Mező, G., Pintér, Á., Kacsuk, P., 2022. The Past, Present and Future of the ELKH Cloud. Inf. Társad. 22, 128. https://doi.org/10.22503/inftars.XXII.2022.2.8

Huang, Y., Ringbauer, H., 2022. hapCon: estimating contamination of ancient genomes by copying from reference haplotypes. Bioinformatics 38, 3768–3777. https://doi.org/10.1093/bioinformatics/btac390

Juhász P., 2015. A honfoglalók létszámának kérdéséhez. Belvedere Meridionale 27, 66–76. https://doi.org/10.14232/belv.2015.3.7

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 356. https://doi.org/10.1186/s12859-014-0356-4

Kovács G., 2022. Középkori és Török kori várak, erődítések Pest megyében. Archaeolingua Foundation, Budapest, Hungary.

Langó, P., 2017. Turulok és Árpádok: Nemzeti emlékezet és koratörténeti emlékek. MTA BTK, Budapest, Hungary.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv Mende, B., Köhler, K., Oross, K., Bondár, M., Marton, T., Osztás, A., Jakucs, J., Paluch, T., Horváth, F., Csengeri, P., Koós, J., Sebők, K., Anders, A., Raczky, P., Regenye, J., Barna, J.P., Fábián, S., Serlegi, G., Toldi, Z., Gyöngyvér Nagy, E., Dani, J., Molnár, E., Pálfi, G., Márk, L., Melegh, B., Bánfai, Z., Domboróczki, L., Fernández-Eraso, J., Antonio Mujika-Alustiza, J., Alonso Fernández, C., Jiménez Echevarría, J., Bollongino, R., Orschiedt, J., Schierhold, K., Meller, H., Cooper, A., Burger, J., Bánffy, E., Alt, K.W., Lalueza-Fox, C., Haak, W., Reich, D., 2017. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature 551, 368–372. https://doi.org/10.1038/nature24476

Maier, R., Patterson, N., Reich, D., 2021. admixtools.

Mallick, S., Reich, D., 2023. The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes Harvard Dataverse. https://doi.org/10.7910/DVN/FFIDCW

Maróti, Z., Neparáczki, E., Schütz, O., Maár, K., Varga, G.I.B., Kovács, B., Kalmár, T., Nyerki, E., Nagy, I., Latinovics, D., Tihanyi, B., Marcsik, A., Pálfi, G., Bernert, Z., Gallina, Z., Horváth, C., Varga, S., Költő, L., Raskó, I., Nagy, P.L., Balogh, C., Zink, A., Maixner, F., Götherström, A., George, R., Szalontai, C., Szenthe, G., Gáll, E., Kiss, A.P., Gulyás, B., Kovacsóczy, B.Ny., Gál, S.S., Tomka, P., Török, T., 2022. The genetic origin of Huns, Avars, and conquering Hungarians. Curr. Biol. 32, 2858-2870.e7. https://doi.org/10.1016/j.cub.2022.04.093

Monroy Kuhn, J.M., Jakobsson, M., Günther, T., 2018. Estimating genetic kin relationships in prehistoric populations. PLOS ONE 13, e0195491. https://doi.org/10.1371/journal.pone.0195491

Olalde, I., Carrión, P., Mikić, I., Rohland, N., Mallick, S., Lazaridis, I., Mah, M., Korać, M., Golubović, S., Petković, S., Miladinović-Radmilović, N., Vulović, D., Alihodžić, T., Ash, A., Baeta, M., Bartík, J., Bedić, Ž., Bilić, M., Bonsall, C., Bunčić, M., Bužanić, D., Carić, M., Čataj, L., Cvetko, M., Drnić, I., Dugonjić, A., Đukić, A., Đukić, K., Farkaš,

Z., Jelínek, P., Jovanovic, M., Kaić, I., Kalafatić, H., Krmpotić, M., Krznar, S., Leleković, T., M. De Pancorbo, M., Matijević, V., Milošević Zakić, B., Osterholtz, A.J., Paige, J.M., Tresić Pavičić, D., Premužić, Z., Rajić Šikanjić, P., Rapan Papeša, A., Paraman, L., Sanader, M., Radovanović, I., Roksandic, M., Šefčáková, A., Stefanović, S., Teschler-Nicola, M., Tončinić, D., Zagorc, B., Callan, K., Candilio, F., Cheronet, O., Fernandes, D., Kearns, A., Lawson, A.M., Mandl, K., Wagner, A., Zalzala, F., Zettl, A., Tomanović, Ž., Keckarević, D., Novak, M., Harper, K., McCormick, M., Pinhasi, R., Grbić, M., Lalueza-Fox, C., Reich, D., 2023. A genetic history of the Balkans from Roman frontier to Slavic migrations. Cell 186, 5472-5485.e9. https://doi.org/10.1016/j.cell.2023.10.018

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient Admixture in Human History. Genetics 192, 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A.L., Pollack, S., Galinsky, K., Chang, C., Gusev, S., Novembre, J., Boursnell, M., Hanna, M., Yu, A., 2017. EIGENSOFT.

Pohl, W., 2018. The Avars: A Steppe Empire in Central Europe, 567–822.

Popli, D., Peyrégne, S., Peter, B.M., 2023. KIN: a method to infer relatedness from low-coverage ancient DNA. Genome Biol. 24, 10. https://doi.org/10.1186/s13059-023-02847-7

Ralf, A., Montiel González, D., Zhong, K., Kayser, M., 2018. Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. Mol. Biol. Evol. 35, 1291–1294. https://doi.org/10.1093/molbev/msy032

Reich, D., 2021. AADR - Allen Ancient DNA Resource [WWW Document]. URL https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data

Révész L., 1996. A Karosi Honfoglalás Kori Temetők - Régészeti adatok a Felső-Tisza-vidék X. századi történetéhez. Miskolc, Hungary.

Ringbauer, H., Huang, Y., Akbari, A., Mallick, S., Patterson, N., Reich, D., 2023. ancIBD - Screening for identity by descent segments in human ancient DNA (preprint). Genetics. https://doi.org/10.1101/2023.03.08.531671

Ringbauer, H., Novembre, J., Steinrücken, M., 2021. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nat. Commun. 12, 5425. https://doi.org/10.1038/s41467-021-25289-w

Rohland, N., Glocke, I., Aximu-Petri, A., Meyer, M., 2018. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. 13, 2447–2461. https://doi.org/10.1038/s41596-018-0050-5

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., Reich, D., 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. B Biol. Sci. 370, 20130624. https://doi.org/10.1098/rstb.2013.0624

Rohland, N., Mallick, S., Mah, M., Maier, R., Patterson, N., Reich, D., 2022. Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. Genome Res. 32, 2068–2078. https://doi.org/10.1101/gr.276728.122

Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., Delaneau, O., 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nat. Genet. 53, 120–126. https://doi.org/10.1038/s41588-020-00756-0

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Sirak, K.A., Fernandes, D.M., Cheronet, O., Novak, M., Gamarra, B., Balassa, T., Bernert, Z., Cséki, A., Dani, J., Gallina, J.Z., Kocsis-Buruzs, G., Kővári, I., László, O., Pap, I., Patay, R., Petkes, Z., Szenthe, G., Szeniczey, T., Hajdu, T., Pinhasi, R., 2017. A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. BioTechniques 62, 283–289. https://doi.org/10.2144/000114558

Szeifert, B., Gerber, D., Csáky, V., Langó, P., Stashenkov, D.A., Khokhlov, A.A., Sitdikov, A.G., Gazimzyanov, I.R., Volkova, E.V., Matveeva, N.P., Zelenkov, A.S., Poshekhonova, O.E., Sleptsova, A.V., Karacharov, K.G., Ilyushina, V.V., Konikov, B.A., Sungatov, F.A., Kolonskikh, A.G., Botalov, S.G., Grudochko, I.V., Komar, O., Egyed, B., Mende, B.G., Türk, A., Szécsényi-Nagy, A., 2022. Tracing genetic connections of ancient Hungarians to the 6th–14th century populations of the Volga-Ural region. Hum. Mol. Genet. 31, 3266–3280. https://doi.org/10.1093/hmg/ddac106

Szőke, B.M., 2021. DIE KAROLINGERZEIT IN PANNONIEN. Verlag des Römisch-Germanischen Zentralmuseums, Mainz.

Szőke, B.M., 2014. The Carolingian age in the Carpathian basin: permanent exhibition of the Hungarian national museum. Hungarian national museum, Budapest.

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. Brill.

Vyas, D.N., Koncz, I., Modi, A., Mende, B.G., Tian, Y., Francalacci, P., Lari, M., Vai, S., Straub, P., Gallina, Z., Szeniczey, T., Hajdu, T., Pejrani Baricco, L., Giostra, C., Radzevičiūtė, R., Hofmanová, Z., Évinger, S., Bernert, Z., Pohl, W., Caramelli, D., Vida, T., Geary, P.J., Veeramah, K.R., 2023. Fine-scale sampling uncovers the complexity of migrations in 5th–6th century Pannonia. Curr. Biol. 33, 3951-3961.e11. https://doi.org/10.1016/j.cub.2023.07.063

Wolfram, H., 1987. Die Geburt Mitteleuropas: Geschichte Österreichs vor seiner Entstehung, 378-907.

Yüncü, E., Işıldak, U., Williams, M.P., Huber, C.D., Vyazov, L.A., Changmai, P., Flegontov, P., 2023. False discovery rates of qpAdm-based screens for genetic admixture (preprint). Genetics. https://doi.org/10.1101/2023.04.25.538339

Acknowledgement

The research presented in this article was funded within the framework of a subproject of the Árpád House Program, initiated by the Hungarian state (The Anthropological and Genetic Image of the Hungarians during the Árpád era, Árpád-house IV. framework program (1109/2018. (III.19.) Gov. decision).

The Research Center for the Humanities of the Hungarian Academy of Sciences contributed to the program with its own concept.

Further funding of the research was provided by the Priority Research Theme proposal of the Eötvös Loránd Research Network (2019-2023 ELKH, 2023- HUN-REN), in the frame of the “Archaeogenomic research of the Etelköz region” project.

Dr. Mihály Molnár (MTA ATOMKI HEKAL, AMS laboratory) is responsible for the radiocarbon analyses in Visegrád. The radiocarbon investigations were carried out in the framework of the OTKA NK 104533 research project of Miklós Takács, which we are grateful for.

On behalf of the bioinfo-baden project we are grateful for the possibility to use HUN-REN Cloud ((Héder et al., 2022); https://science-cloud.hu/) which helped us achieve the results published in this paper.

We are also grateful to the following archaeologists for their contribution to the project: Péter Tomka (for his advices about Himod-Káposztásföldek), Judit Kodolányi (Visegrád-Sibrik hill I), Gergely Buzás and Szabina Merva (Visegrád-Sibrik hill II), and István Kováts, Péter Gróf (Visegrád- Széchenyi street), Ágnes Ritoók (Zalavár), as well as anthropologists Tamás Szeniczey and Krisztián Kiss. We would like to thank for the invaluable comments on the paper to Péter Langó, Tivadar Vida, István Koncz, Balázs Gyuris and Levente Samu.

Author contribution

Designed the study: BGM., MBSz., ASz-N.
Collected data: CsL, SÉ, MBSz., ZsP., FSz, PR, BGM.
Laboratory analyses: VCs, BSz,
Performed analyses: DG, KJ, GyM, BGy, VCs, BSz, NB
Wrote the paper: DG, ASz-N
Supervised the manuscript: ASz-N, MBSz

Ethics declarations

The authors declare that they had requested and got permission from the stakeholders, excavator, and processor anthropologists and archaeologists for the destructive ancient DNA analyses of the human remains presented in this study.

Competing interests

The authors declare no competing interests.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The newly produced sequence data is deposited in the European Nucleotide Archive (ENA) with the following accession number: PRJEB75908. We also report on samples that have not yielded sufficient ancient DNA in Table S1.

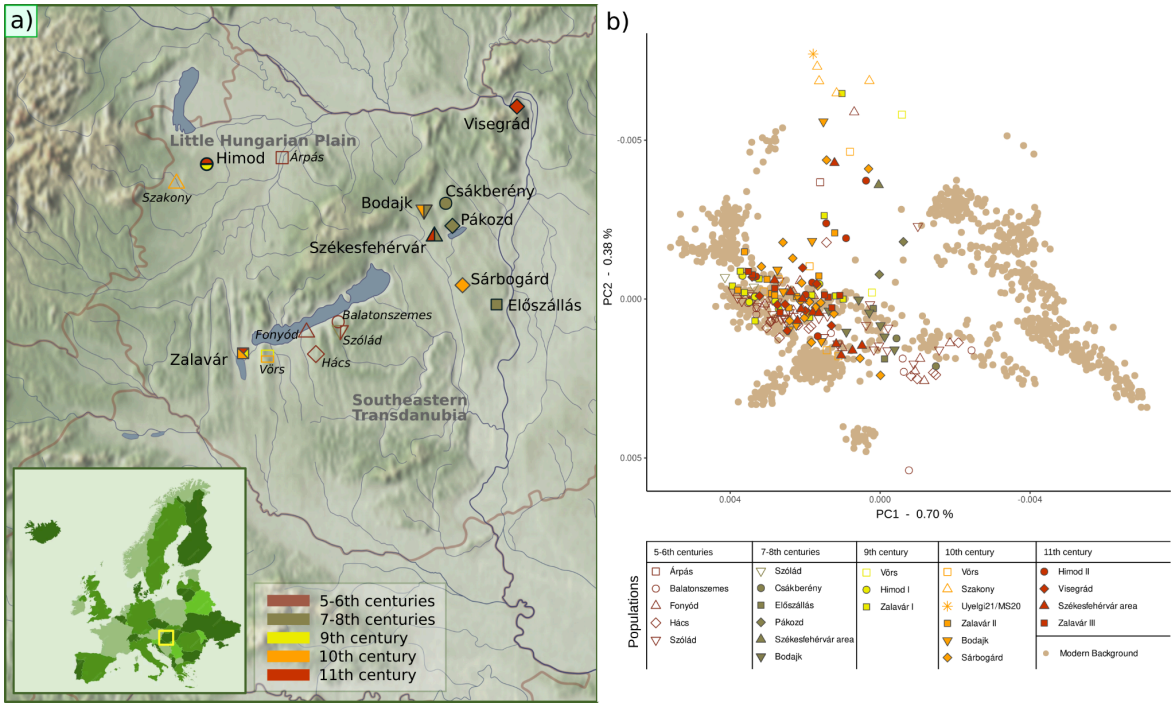

Figure 1. Sites and general genomic composition of the samples originating from Transdanubia. *a) Discussed sites in Transdanubia, where italic labels and empty symbols stand for previously published datasets (Amorim et al., 2018; Maróti et al., 2022; Vyas et al., 2023) b) PCA of Transdanubia between the 5th to 11th centuries CE. Ancient published (empty symbols) and newly sequenced genomes are projected onto modern West Eurasian genomes.*

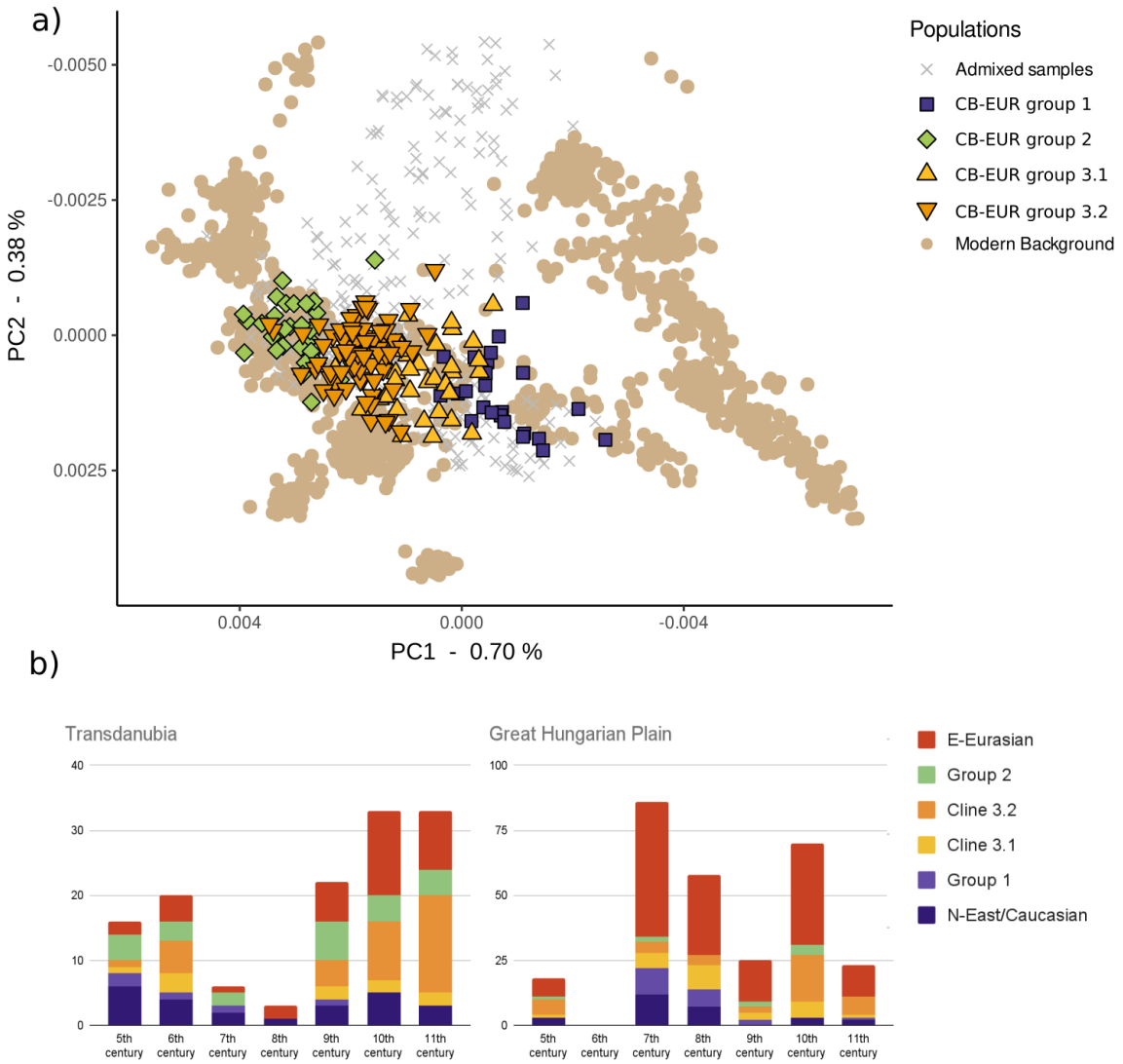

Figure 2. Distribution of European genetic groups. *a) PCA representation of f_4-cluster based classification of individuals bearing only European ancestry. Further tests (Supplementary Information section 3.2) revealed that these groups represent sections of a regional genetic cline, where groups 1 and 2 represent the two extremes and cline 3 the genetic cline between these. b) Chronological and geographical distribution of the genetic groups/clines. Until a homogenisation process in the 10th century CE, clear differences between the two macroregions can be observed.*

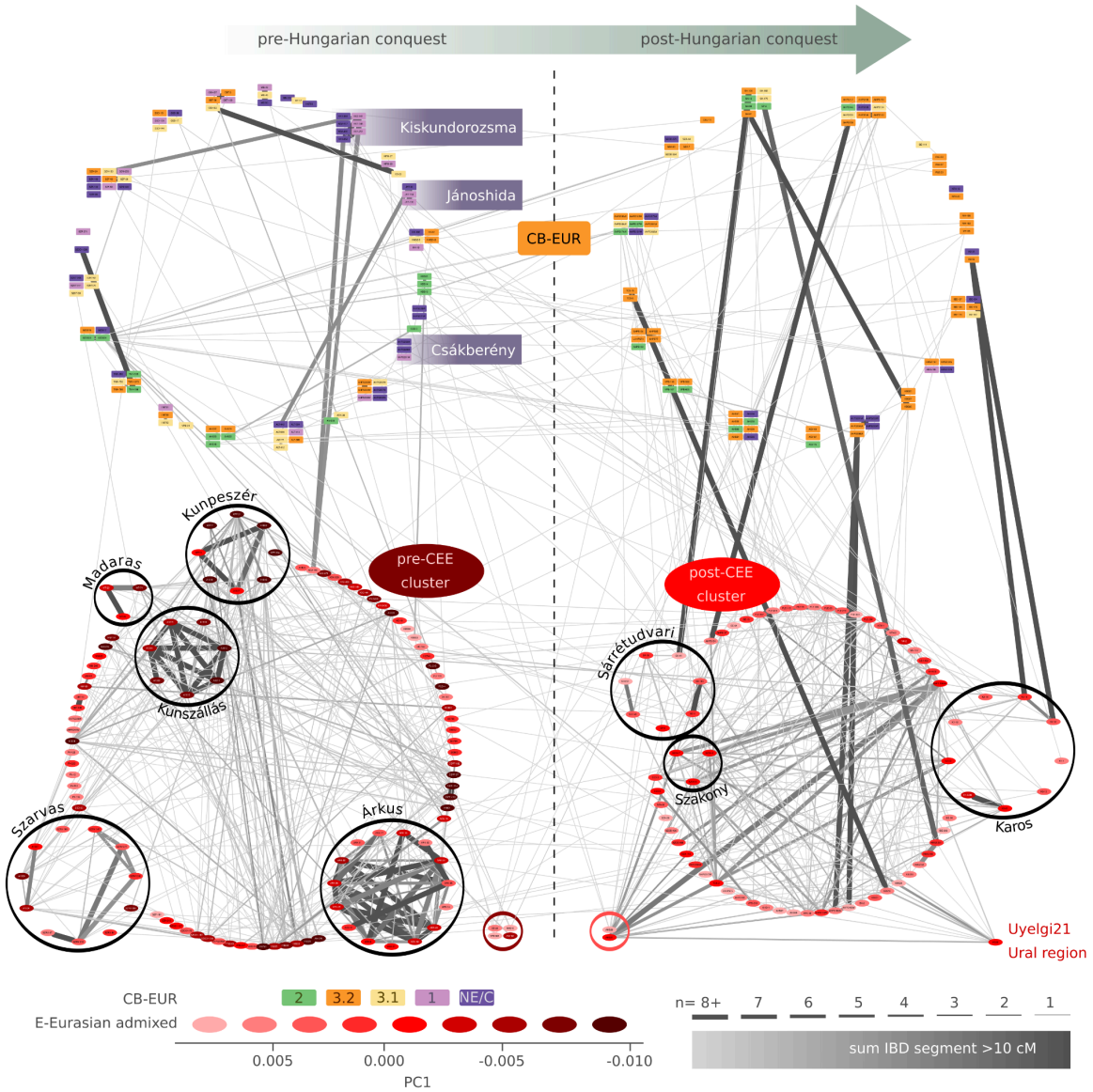

Figure 3. IBD segment sharing networks of the Carpathian Basin populations in the 7th-11th centuries CE. *The colouration of the individuals in CB-EUR corresponds to their European genetic groups/clines (PC1 relative positions are shown at the bottom), the samples are grouped by sites (Table S5). Ovals in the lower clusters represent individuals with significant East-Eurasian ancestry and colour depth reflects their PC1 positions (PC1 relative positions are shown at the bottom), they are also sorted by sites. Sites mentioned in the text are circled in black. The width of the edges corresponds to the number of >10 cM IBD segments and the colour depth signals the summarised length shared in cM. Three groups are highlighted due to their homogeneous CB-EUR grouping. Individuals of different chronological and East-Eurasian cluster classification are separated and circled in red at the bottom of the graph. For individual classifications, see Table S5, for IBD connections Table S6.*

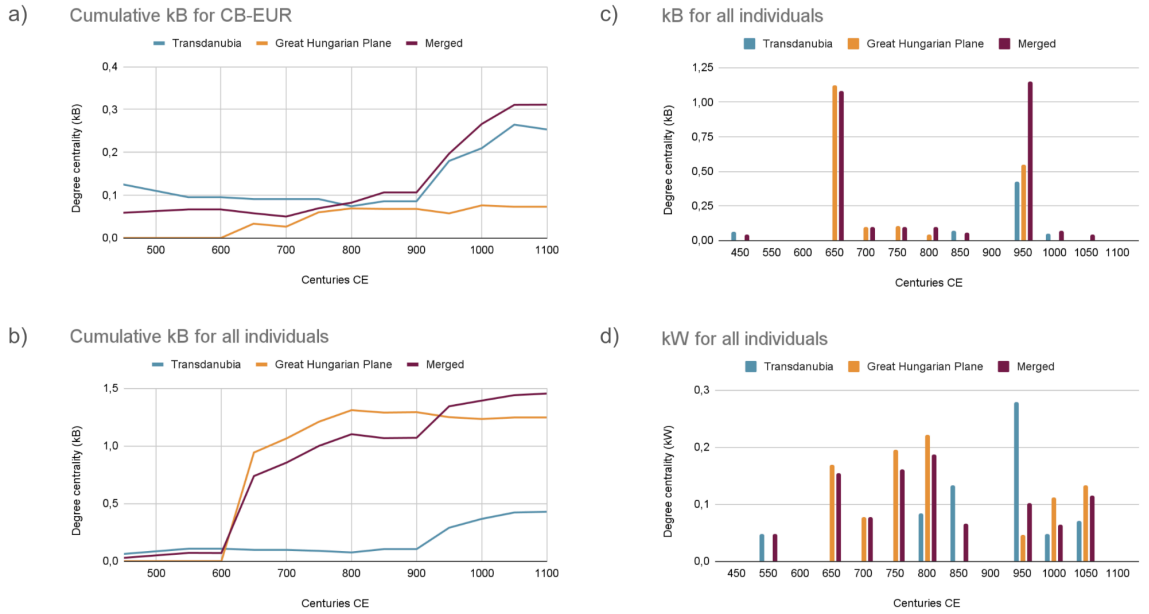

Figure 4. Intra- and intergroup tendencies in IBD segment sharing patterns. *a) Cumulative connections between sites (k_B) regarding the CB-EUR cluster in a chronological order signal movement within and to Transdanubia in the 10th century CE but not within GHP. b) The same analysis when including all individuals signals a movement and admixture between sites in GHP at the 7th century CE, which likely can be connected to the settlement and exogamous social structure of the Avars. c) The non-cumulative approach of connections between sites signals the previously observed two waves of movements but also shows little to no migration between these periods. d) Within site connections (k_W) reveal degrading biological networks between sites separating the two waves of migrations.*

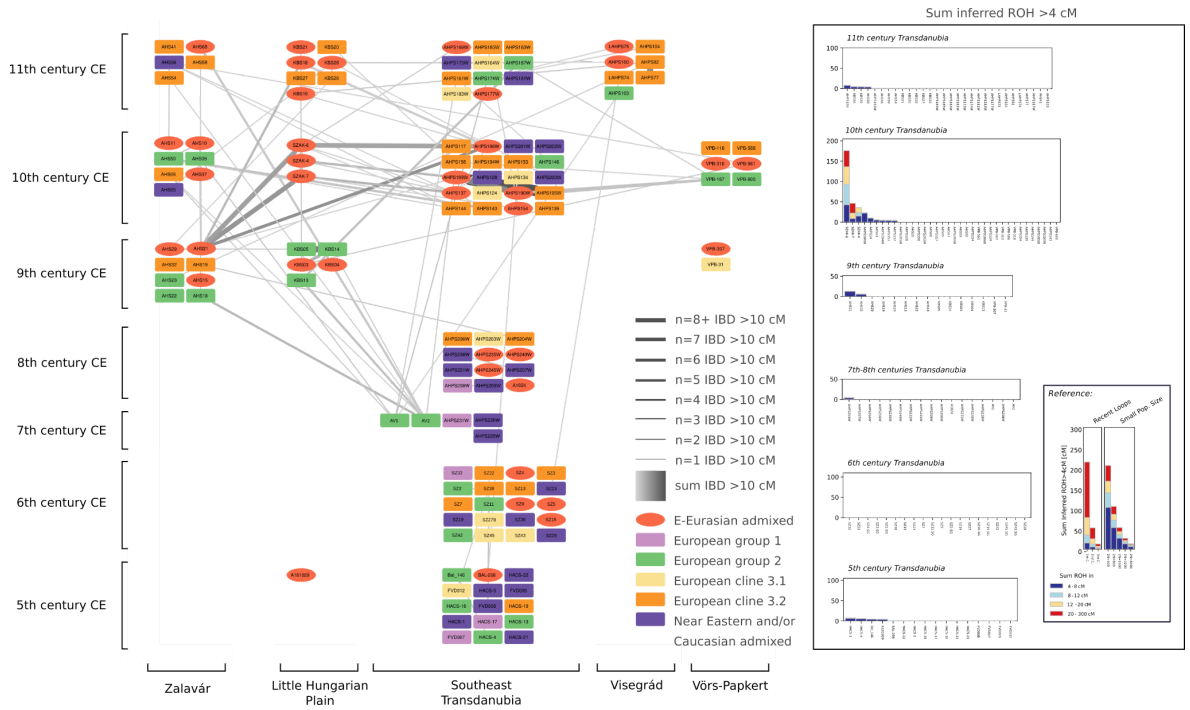

Figure 5. IBD patterns across Transdanubian sites in the 5th-11th centuries and their corresponding ROH patterns. *Only connections between 10 and 1400 cM are visualised, in order to emphasise population level connections instead of 1-2nd degree connections.*

Table 1. Summary of the sampling and analyses carried out in this study

Site and phase	Ecoregion	Century CE	Sample number	Shotgun genome	MtDNA	Y-STR
Bodajk-Homoki dűlő	SE-Transdanubia	7-9	19	7	7	-
Bodajk-Proletárföldek	SE-Transdanubia	10	14	8	8	-
Csákberény-Arató szérű	SE-Transdanubia	9	8	1	1	-
Csákberény-Orondpuszta	SE-Transdanubia	6-7	9	2	2	-
Előszállás-Bajcsihegy	SE-Transdanubia	7-8	6	2	2	-
Himod-Káposztásföldek I	Little Hungarian Plain	9	14	7	14	2
Himod-Káposztásföldek II	Little Hungarian Plain	10-11	14	7	14	2
Pákozd-Börgöndpuszta	SE-Transdanubia	8	11	2	2	-
Sárbogárd-Tringer tanya	SE-Transdanubia	10	42	18	40	-
Székesfehérvár-Rádiótelep	SE-Transdanubia	10-11	17	6	6	-
Székesfehérvár-Sárkeresztúri út	SE-Transdanubia	10-11	6	4	4	-
Székesfehérvár-Sóstó	SE-Transdanubia	8	5	1	1	-
Székesfehérvár-Homokbánya	SE-Transdanubia	10-11	3	1	1	-
Székesfehérvár-Táci út	SE-Transdanubia	10-11	3	1	1	-
Visegrád-Sibrik domb I	Visegrád	11	35	6	25	-
Visegrád-Sibrik domb II	Visegrád	11	3	-	2	-

Visegrád-Széchenyi utca 25.	Visegrád	11	5	3	5	-
Zalavár-Vársziget I	Zalavár	9	25	11	25	24
Zalavár-Vársziget II	Zalavár	10	22	8	21	9
Zalavár-Vársziget III	Zalavár	11-12	25	8	25	19
Uyelgi, Russia	Trans-Ural region	8-10	1	1	1*	1*
SUM			295	103	206	56

*published in Csáky et al. 2020