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[1,6,9,10], 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

Abstract

During the Early Medieval period, the Carpathian Basin witnessed significant demographic shifts, notably under the 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. By employing identity-by-descent (IBD) segment sharing networks, this research offers detailed insights into the population structure and dynamics of the region from the 5th to the 11th centuries CE, with specific focus on certain microregions. Our evaluations reveal spatially different histories in Transdanubia even between communities in close geographical proximity, highlighting the importance of dense sampling and analyses. Our findings highlight extensive homogenisation and reorganization processes, discontinuities between Hun, Avar, and Conquest period immigrant groups, alongside the spread and integration of ancestry related to the Hungarian conquerors.

Teaser

Genetic formation of medieval 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;
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 - post-Conquest East-Eurasian, genetic cluster of individuals with East-Eurasian ancestry chronologically following the Hungarian conquest;
pre-CEE - pre-Conquest East-Eurasian, 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-Vársziget site.

Introduction

The population history of the Migration period in Central Europe was marked by broad-scale complexity, influenced by various Eastern and Western cultural and biological impacts, as documented in written sources and archaeological studies[1], [2], [3], [4], [5], [6], [7]. In recent years, the Carpathian Basin (including the territory of today's Hungary) has been the target for many archaeogenomic studies[8], [9], [10], [11] shedding step-by-step light on population events of this period.

Between the fall of the Roman Empire in Pannonia (433 CE) and the arrival of the Avars in 568 CE, the Carpathian Basin was home to transient groups and polities, such as the Suebs, Goths, Lombards, and Gepids, with diverse ethnic and cultural backgrounds. During this period, the Avar elite of Northeast-Eurasian origin[9], [10], [12], along with various East-European groups, settled in the region[4]. According to the written and archaeological sources, the population of the Avar Khaganate was extremely culturally diverse[5], [13]. A 20-year 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 the spread of its cultural influence and strategic integration of the corresponding regions during the 9th century CE. This involved the establishment of smaller, isolated counties as administrative units and a deliberate effort to Christianize the people of the newly acquired territories[5].

The power center 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, where the elites and their servants were being buried. Around 880 CE, the settlement became a royal center 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 new arrivals overran Transdanubia in the 10th century CE and settled across the region, presumably displacing only the ruling class of the former population. Zalavár remained a prominent settlement and was recognized as the capital of both Somogy and Zala counties, which still exist today. The population density increased, and a Benedictine church was consecrated in 1019 CE, with a cemetery around it used in 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.

The entire 10th century is referred to as the Conquest period. During this time, all previous administrative structures were disregarded, and new centers were established[5]. 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 the GHP[9], [10], 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 a more comprehensive genetic dataset 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 to establish the 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 to analyze 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.

Overall, 294 samples were collected from various sites in Transdanubia (see Fig. 1.a, Table 1 and Supplementary Information section S1, Table S1), with special attention given to Zalavár, the Székesfehérvár area, and Visegrád (Visegrád-Sibrik hill and Visegrád-Széchenyi street 25 sites), as each represent key aspects of the studied historical transitions. The Zalavár site, known as the greatest center under the name *Mosaburg* during the Carolingian era, was sampled across three consecutive archaeological horizons: the 9th century (ZVI), the 10th century (ZVII), and the 11th century (ZVIII). Székesfehérvár, a newly formed settlement in the 10th century CE, became the royal center of the Hungarian Kingdom. The population context for the preceding centuries is represented through sampling of neighboring sites (Supplementary Information section 1). Himod-Káposztás (hereafter Himod) represents a transitional smaller rural settlement, which was sampled from two horizons (9th and 10th-11th centuries CE), while Visegrád was a newly formed religious and political center 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, resulting in genomes with ~0.019-2.14× average genomic coverage (mean ~0.86×). The number of detected SNPs from the 1240k panel[14] ranges from 18,875 to 1,008,685 (mean ~563,152), enabling a deep analysis of the dataset. In addition, 184 mitochondrial genomes via mitogenome capture and 56 Y-chromosomal STR profiles were created for samples from the Zalavár, Himod, Visegrád and Sárbogárd sites (Supplementary Information section 2, Table S1). Sample Uyelgi21, under the ID MS20 from the 8th-10th century CE Trans-Uralic Uyelgi site[15], was resequenced for its importance in Hungarian

prehistory to 1.89× average genomic coverage (952,072 genomic SNP on 1240k, Table S1), and used as a proxy to distinguish Northeast-Eurasian and Western-Siberian genetic ancestries and connection systems in our analyses. For further sample information, see Supplementary section S1. In the following sections, we co-analyzed our dataset with aDNA samples from the region of present-day Hungary, dated between the 5th and 11th centuries CE[8], [9], [10], [11]. For *f*-statistics, we also included samples from Antonio et al.[16]. 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 a modern West-Eurasian background (Figures 1.b and S6-S10), and the K=6 unsupervised *ADMIXTURE* analysis (Fig. S11, Table S2) show large diversity among the studied samples from Transdanubia (5th-12th centuries CE). This diversity is also reflected in the uniparental makeup that represents a wide range of Eurasian macrohaplogroups at both the mitochondrial and Y-chromosomal levels (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[17] 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. This is also reflected by PCA positions (Fig. 2a) and *ADMIXTURE* proportions (Fig. S13). After performing further f_4 tests (Supplementary Information section 3.2), we found that these groups form a genetic cline in the region, distributed roughly in a North-South axis (Fig. 2a, Table S5). Groups 1 and 2 represent the extremes, while Group 3 is generally a mixture or cline between them. Therefore, we refer to it as European Cline 3 in the following analyses. This description is a refined version of the Carpathian Basin European diversity characterized by Maróti et al.[10]. We also discovered that Near-Eastern ancestry is strongly correlated with Group 1 and Cline 3.1 (Supplementary Information section 3.2). The genomic makeup of Group 1 and Cline 3.1, along with their PCA positions and uniparental markers, indicates a strong Late Antique and/or Balkan influence in the studied region, as also observed by Olalde et al.[18]. This suggests similar impacts in the territory of present-day 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 partially described by Vyas et al.[11]. Our f_4 grouping results were also confirmed by subsequent IBD analyses (Supplementary Information section 3.2, Fig. S21). We refer to individuals with exclusively European, Near-Eastern, and Caucasian ancestry as CB-EUR hereafter, as the latter two often co-occur and do not show specific ingroup separation in regional IBD analyses (see below). Based on Olalde et al.[18], these ancestries appeared in the Balkans by the end of the Roman period and are considered to be part of the autochthonous population makeup of the studied era.

The chronological and geographical distribution (Fig. 2b) of these genetic groups/clines and individuals with non-European ancestry varies over time and between Transdanubia and the GHP until the 10th century CE (Supplementary Information section 3.2, Fig. S16). Roughly half of all investigated samples possessed some non-European ancestry (tested by f_4-statistics), with the prevalence of East-Eurasian ancestry being twice as high in the GHP compared to Transdanubia (Fig. 2b, Table S4). It is important to note that f_4-statistics may not detect trace amounts of ancestries. Therefore, minor East-Eurasian ancestry may be present in the CB-EUR individuals, but this did not distort further analyses.

Genomic haplotype segment sharing within the Carpathian Basin

Due to the methodological limitation of the allele frequency based methods[19], we also applied IBD analysis using *ancIBD*[20] on high-quality samples after *GLIMPSE* phasing and imputation[21]. For details, see Materials and Methods section, and for the results, see Supplementary Information section 4 and Table S6.

Text box 1: Benefits and drawbacks of IBD-based analyses.

1) The accuracy of phasing and imputation, which highly depends on the background database, genome coverage, and data type (capture or shotgun), significantly affects the correct recovery of shared IBD segments according to estimates by Ringbauer et al.[20]. Although ancIBD robustly detects IBD longer than 8 cM[20], we restricted analyses to >10 cM segments in order to minimize false signals.
2) Unlike allele-frequency-based population genetic analyses, the IBD-based analyses do not inherently require the exclusion of close relatives. However, for certain IBD-based analyses (as described in the Methods section), we opted to exclude relatives to avoid potential distortions in the results.
3) Shared IBD segments tend to shorten, and 10 cM segments disappear over a few hundred years, which is a phenomenon that must be considered when interpreting the data. Therefore, the lack of shared IBD segments between two individuals or clusters/groups, especially with a wider chronological gap, does not necessarily indicate a lack of genetic connections, but could still reflect no or limited shared genetic ancestry. Additionally, effective population size, community organization and sample number also influence the pattern of observed IBD sharing.
4) Recovering shared IBD segments between individuals and clusters/groups does not directly reflect the ancestry composition of the studied populations. Therefore this evidence needs to be discussed separately. Shared IBD segments between two individuals could either reflect a direct relationship between them or heritage from their common ancestry.

We defined 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 an edge-weighted spring embedded layout on the undirected graph, two major clusters are formed (Fig. S19) by individuals with predominant East-Eurasian ancestry. This largely overlaps with chronology – with the clusters separated by the event of the Hungarian

Conquest approximately during the 890's CE. The individual Uyelgi21, associated with the late Kushnarenkovo culture in the Southern Urals (8th-10th centuries CE), shows extensive connections to post-Conquest individuals with East-Eurasian ancestry (post-CEE) and only one 10 cM connection to a pre-Conquest cluster of individuals with East-Eurasian ancestry (pre-CEE), reinforcing the separation of the two clusters (Fig. 3). Individuals of CB-EUR exhibit a rather homogeneously scattered network, likely due to a larger population size (see below). Next, we redefined these East-Eurasian ancestry clusters based on their 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 generally align, we identified some exceptions (Table S5, detailed in the Discussion).

Most of the post-CEE cluster individuals are situated on a different cline on the Eurasian PCA compared to the pre-CEE cluster individuals. Many are positioned towards and almost overlapping with Uyelgi21 (Fig. S12), consistent with its extensive IBD connections to them. Despite several 10-20 cM connections (Table S6.1), we find predominant discontinuity between pre- and post-CEE clusters as well (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 (Table S4.3). This analysis aimed to determine whether the sporadic connections between pre- and post-CEE clusters are through CB-EUR individuals, local East-Eurasian admixture, or common East-Eurasian ancestry. Accordingly, we observed the latter only through a few connections (Table S6.1). These connections accumulate among post-CEE individuals (with one exception, see Discussion) from the second half of the 10th century CE in some scattered areas, indicating sporadic and regionally varied survival of these lineages into the post-CEE cluster, suggesting local admixture rather than common eastern ancestry.

The IBD connections of the pre-CEE cluster to CB-EUR are much 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). Conversely, the average within-cluster connections for pre-CEE (~3.16) and post-CEE (~2.55) are rather similar, and are much higher, compared to the pre-CB-EUR (contemporaneous CB-EUR individuals to pre-CEE, ~0.38) and post-CB-EUR (~0.30) (for individual classification, see Table S5). 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-cluster, except for one sample (KDA-188, Kiskundorozsma site), and are all restricted to the same site. Conversely, while the post-CEE cluster connections are also mostly restricted to the same site, they primarily occur between CB-EUR and post-CEE individuals. Interestingly, both post-CEE males and females were involved in this admixture with CB-EUR individuals, likely in equal proportion. This conclusion is based on the distribution of presumably Eastern Y chromosomal (C2, Q1, N1a, R1a-Z93) and mtDNA haplogroups (A, B, C, D, G, M, N1a1a1a1a, N9, R, U4d2, X2f) found in the post-CEE individuals, alongside West-Eurasian uniparental markers (Table S10)[10], [15], [22]. 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, consistent with the observations of Gnecchi-Ruscone et al.[23] and our f_4-results (Supplementary Information section 3.2 and Fig. S16).

The IBD results, particularly the above outlined network density differences between clusters, become more conspicuous when considering the effective population size (N_e) of the pre-CEE cluster compared to the post-CEE cluster, calculated with *hapROH* (see Methods). We calculated and compared the N_e of the pre- and post-CEE clusters to their contemporaneous subset of CB-EUR individuals (Table S8). Estimates on actual population size are problematic due to the lack of solid historical or anthropological demographic data for the period, but approximations were made nonetheless[24]. Our analysis suggests that around 20.82% of the population of the Avar period belonged to the pre-CEE cluster, whereas only approximately 13.01% of the population from the 10th-11th centuries CE belonged to the post-CEE cluster. This explains the IBD network density differences between the clusters (*i.e.* a smaller population has more internal IBD sharing), and aligns with historical estimates of 500k to 2 million individuals for the Carpathian Basin and 50k to 150k individuals for Hungarian conquerors (which may overlap with post-CEE, see Discussion)[24]. Additionally, the N_e estimates for CB-EUR indicate a small but definite increase towards the end of the millennium (Table S8). We note that N_e might be heavily influenced by societal customs and marriage patterns, therefore it is not necessarily indicative for census population sizes.

The IBD connections of CB-EUR (Fig. S21) reflect the f_4-based classification and also suggest that different sections of the CB-EUR cline likely consist of distinct communities and groups. For example, those belonging to European Group 2 show discontinuity before and after the 7th century CE (Fig. S22), indicating the presence of different groups that share a similar regional genetic makeup.

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

To further understand the social and regional dynamics of communities between the 5th and 11th centuries CE, we analyzed trends in community networks based on IBD data. We explored weighted degree centrality (i.e., the number of connections, denoted as k, calculated separately for between-site (k_B) and within-site (k_W) interactions) using different weighting methods (see Methods for details and Table S6.2 and Fig. 4 for results). We calculated k_B and k_W values separately for Transdanubia and the GHP, as well as for the combination of these two macroregions, denoted as 'Merged' in Fig. 4. The k values were also computed for each time interval (*i.e.* centuries), as shown in Figures 4c and 4d. This approach allows us to capture both spatial and temporal differences in the network structure. Furthermore, we were particularly interested in k values of the CB-EUR cluster alone, as it might behave differently compared to the broader dataset where non-local elements are overrepresented. Due to the low number of CB-EUR connections, we developed a cumulative approach for k_B. Although we used a different formula with two separate weighting methods (see Methods), the theory is similar to the conventional approach, as it involves using a weighted number of connections and adding samples from previous batches to each subsequent batch. The rationale behind this approach is to detect connections that might otherwise be masked by the

narrow time window. Using our formula, this method highlights local trends that would otherwise remain hidden (Figures 4a and b).

Cumulative k_B trends differ remarkably when calculating these for only CB-EUR versus all individuals, including those with East-Eurasian ancestry (Figures 4a and 4b), which points to vastly different social organizations of local (CB-EUR) communities compared to those who possess East-Eurasian ancestry. However, certain commonalities also emerge: 1) The 'Merged' values, which signify connection between the Transdanubian and the GHP communities, shows the same relative trend in both cases. Accordingly, Transdanubia and the GHP remained relatively separated until the turn of the 9th-10th century CE, when, likely fueled by the Hungarian conquest, contacts between these regions increased. 2) The GHP communities, contrary to the Transdanubian communities, show no change in k_B values during this turning point, pointing to movements from the GHP to Transdanubia. This trend is also much more pronounced for the Transdanubian CB-EUR group when using one of our alternate weighting methods (Methods, Fig. S23). The sudden increase of intersite connections in the 7th century CE onwards in the GHP communities are in line with the observations of Gnecchi-Ruscone et al.[23], who suggested sedentary lifestyle along with female exogamy within the eastern-origin strata of the Avar period.

The conventionally calculated k_B (Fig. 4c) revealed two major peaks in the 7th and 10th centuries, indicating two waves of intensive biological contact between sites, and stability in between. The first wave correlates with the changing Avar society in the 7th century CE, while the second wave aligns with the reorganizations fueled by the Hungarian conquest at the turn of the 9th-10th centuries CE. The stable lifestyle in rather smaller communities is reflected by some 7th century CE sites, such as Kunpeszér and Kunszállás from Maróti et al.[10] and Gnecchi-Ruscone et al.[9], and even by the few samples from Madaras (Fig. 3), a large burial site of the 7th century CE, which show high degree of within-site connections, regardless of biological sex. This observation further diversifies previous findings about female exogamy during this period[23]. In addition to that, the data also show a notable rise of between-site connections at the turn of the 8th-9th century CE in the GHP, indicating some degree of population movement or change during that time. In Fig. 4d, the k_W values complement the k_B in Fig. 4c. There are several notable features in the joint interpretation of Figures 4c and 4d. The first is the increased contacts around the turn of the 8th-9th centuries CE in the GHP compared to the previous century, which may reflect events related to the changing political environment described in the introduction, although other scenarios are also possible. Another notable feature is the high k_W value for 10th century CE Transdanubia, especially in comparison to that of the GHP. This phenomenon could result from chronological assignment bias, although the sample number is the highest in this 10th century CE batch (103 overall, 33 for Transdanubia), and the k_B values of the same batch are synchronous with those of the GHP, making this scenario unlikely. If the observed k_W is a true feature of the dataset, it may signal a few generations' delay of movements in Transdanubia compared to the GHP. This scenario correlates with archaeological-historical data that the Hungarian conquest was a sequence of events in the Carpathian Basin from the eastern to western territories[25]. On the other hand, this phenomenon could also be the result of

different sampling strategies between the Transdanubian and GHP sites, as in the study by Maróti et al.[10] many characteristically military-related cemeteries were sampled, which have a different organization and IBD pattern compared to biological community-based community burial sites. After the Hungarian conquest, k_W values tend to slowly increase in the following centuries in both regions, pointing to a more stable and established community system.

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. A common element between regions is the proportional decline of European genetic Group 1 and Near-Eastern ancestry, as well as the proportional enrichment of Cline 3 (*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. S22). This observation underscores that PCA and evolutionary-oriented techniques, such as *ADMIXTURE* and *f*-statistics based methods, are limited in their ability to correctly characterize populations in a frequently intermixing milieu. Therefore, conclusions drawn from such results must be approached with great caution. However, certain trends should not be overlooked. For example, in the pre-Conquest period, the presence of a remnant Romanised population around Csákberény, based on archaeological findings (Supplementary Information section 1.8), is paralleled by the Group 1/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 individuals, and the GHP sites Kiskundorozsma and Jánoshida[10] with Group 1 individuals, although these are less corroborated by archaeological background (Fig. 3). This genetic isolation is, however, mitigated by connections to pre-CEE individuals, such as one in Kiskundorozsma.

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

Recent researches have highlighted the predominance of East-Eurasian ancestry among the early Avar Khaganate's elite, indicating these groups as a significant genetic source of immigrants (Fig. S12)[9], [10], [23]. 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 the ethnic identities and genetic categories used do not correspond directly. Similarly, the post-CEE cluster can be strongly associated with the Hungarian conquerors and their descendants. The remarkable discontinuity between pre- and post-CEE clusters is nuanced by the sporadic IBD connections, which almost exclusively accumulate in cemeteries from the second half of the 10th century CE (Table S6.1), indicating limited local survival of pre-CEE cluster lineages. The only systematic 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[26]. Additionally, contrary to the suggested continuity of 5th century CE Hun period East-Eurasian genetic ancestry into the Conqueror cluster by Maróti et al.[10], our data indicate an absence or loss of this ancestry among post-CEE individuals in the studied region (Fig. S24). However, the greater time gap and the limited number of Hun period samples in the Carpathian Basin may obscure distant and sporadic genetic connections.

The IBD connections of the Hungarian conquerors to CB-EUR demonstrate an immediate mingling 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. This difference is also highlighted by the results of f_4-statistics, which suggest different West-Eurasian ancestry for pre- and post-CEE clusters (Supplementary Information section 3.2, Fig. S16). These results, however, can be interpreted in various ways. The genetic isolation of the Avar elite previously demonstrated by studies[9], [12], [23] is now corroborated by our IBD analyses. This highlights the distinct social organizations of the communities belonging to the pre- and post-CEE clusters. Our findings provide a nuanced perspective on the observations of Gnecchi-Ruscone et al.[23], regarding extended exogamy as a general residence pattern of the Avar society. Based on the results summarized in Fig. 4, Avar communities show two waves of between-site connections (likely driven by female exogamy) first in the 7th century CE and again at the turn of the 8th-9th century CE. This observation is further supported by the high within-site kinship-level connections of late Avar period sites, such as Árkus and Szarvas (Fig. 3), irrespective of biological sex. Notably, at the sites investigated by Gnecchi-Ruscone et al.[23], many females did not connect to the studied pedigrees, and many pedigree-forming females are missing from the family trees. In a broader multi-site (macroregional) model, our findings depict a different social structure. However, we cannot rule out exogamy from closely neighboring villages during periods where we observe a drop in k_B values. This idea is also supported by the strontium isotope results from Gnecchi-Ruscone et al.[23], which did not indicate large-scale movement between distant villages. The relative isolation found among the Avars (Fig. 3) may explain their eventual disappearance and the subsequent genetic and cultural succession by the Hungarians. Furthermore, these results suggest that the West-Eurasian (likely Eastern-European, Steppe and/or Caucasian) admixture in the pre-CEE cluster primarily originated outside the Carpathian Basin, consistent with previous observations[23]. This

indicates a prevailing West-Eurasian genetic substrate during the Avar period, which is currently traceable only through its 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). Two individuals from the early 7th century occupation of Szólád (AV1 and AV2, first-degree relatives) display extensive connections to individuals from most of the studied sites, with the most notable connections being to Zalavár (*Mosaburg*). Recovered from deposit fillings of a Langobard period grave without any artifacts[8], these individuals nonetheless indicate the presence of a genetically successful (and likely large) substrate in the region, probably arriving around the 7th century CE. Given this genetic substrate's presence centuries before *Mosaburg* was founded, we cannot confirm the written sources that claim the entire settlement's population originated outside the Carpathian Basin, suggesting possible exaggerations in historical records. Nevertheless, the presence of mtDNA haplogroup C1e (AHP20), previously only found in Icelandic populations[27], and the Y-chromosomal haplogroups I1-Y6357 (AHP16) and R1a-FGC11904 (AHP18), primarily of Northwest-European origin (Yfull.org) suggest some foreign connections in the Carolingian horizon of Zalavár. Although the sample size is small relative to the presumed large populations of the past, the Zalavár sample set demonstrates some continuity from the 9th to the 11th centuries CE.

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. Connections between the pre- and post-Conquest phases (I-II) of the Himod site are sporadic, and differences in admixture proportions (Fig. S11) and uniparental makeup suggest limited genetic continuity between the horizons. Instead, these results rather reflect new Hungarian occupation, with incoming new genetic components connected to the GHP and the post-CEE cluster, 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, corroborating written sources that describe it as a newly formed community[28]. However, future comparative IBD analyses with other European areas are required to fully understand the composition of this society. In the thoroughly analyzed Székesfehérvár area, we observe a different pattern compared to Zalavár. A Late Antique-type population persisted in this region, and could even be found around Sárbogárd in the 10th century CE. However, the appearance of Eastern elements after the Hungarian conquest is much more prominent in the area. Despite the larger number of samples from southeastern Transdanubia, signs of local succession are less pronounced in this area. This is consistent with expectations of community restructuring after the Hungarian conquest, as described by the IBD degree centrality changes throughout the centuries. Overall, the microregional evaluation reveals spatially different histories even between communities in close geographical proximity, highlighting the importance of

prudently integrating 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 the sites throughout the 5th-11th centuries reflect both Avar and Hungarian conqueror occupations 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. However, the Hungarian conquest fueled intensive population movements and admixtures within the Carpathian Basin, consistent with historical records [5]. This process is further evidenced by the occurrence of relatives in geographically distant cemeteries after the Hungarian conquest. For instance, individuals KOZ018 (Himod, Transdanubia, 10th-11th century CE) and TCS-18 (Tiszanána, GHP, 10th century CE) are a father-son pair, while KOZ026 (Himod, Transdanubia, 10th-11th century CE) and SH-81 (Sárrétudvari, GHP, 10th century CE) are 3rd-degree relatives (Table S6). We observed a potential delay in this process in Transdanubia compared to the GHP, which underscores another difference between these two microregions, and highlights the importance of considering ecological barriers, such as the Danube River, in reconstructing past population events.

We also found evidence for the pre-conquest presence of individuals with Uralic ancestry. In phase ZVI (dated to 850-890), the adult male AHP21 shows strong connections to multiple individuals in the post-CEE cluster. This individual's genomic composition is highly similar to Uyelgi21 with strong IBD connections (Table S6 and Fig. S12), and is a 5th-degree relative of adult male KeF1-10936 from Kenézlő-Fazekaszug (GHP), whose dating can only be approximated to 900-960 CE. Additionally, AHP29 from the Carolingian era (ZVI) also shares IBD segments with the Conqueror cluster, although its connections are much weaker compared to those of AHP21's. Both burials firmly predate the Hungarian conquest based on archaeological stratigraphy (Supplementary Information section 1). These findings signal sporadic appearances of Hungarians in Transdanubia before 890 CE (new radiocarbon date of AHP21 is in Table S1, stratigraphic date is in Supplementary Information section 1). These individuals likely served as mercenaries at the court of Priwina, according to 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 the database data with ours, as well as the methods we used, provides enough statistical strength to the statements of this study. However, since we did not include comparative datasets from other countries—focusing instead on the local characterization of population events—we acknowledge that certain questions remain open. These include the ancestry shift within

CB-EUR Group 2, the origins of the Late Antique-type population, and a more detailed characterization of the post-CEE cluster. These issues could be more fully addressed through exhaustive comparisons with data from other populations in future research. We believe that further sampling from East-Central Europe is required to reveal such intricate questions, and we aim to do so for subsequent studies.

Materials and Methods

Ancient DNA laboratory work

Samples were 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.[22]. Bone powders were produced based on Lipson et al.[29] and Sirak et al.[30], DNA extractions were performed following the methods described in Dabney et al.[31] and Rohland et al.[32], and the double stranded DNA libraries were created using half-UDG treatment[33]. DNA extraction and library preparation were performed either manually or automatically with Biomek i5 Liquid Handling Workstations. Initial screenings were performed on an inhouse Illumina MiSeq platform, and eligible samples were then deeper sequenced on high throughput Illumina platforms (Applied Biotechnologies and Novogene), for all results, see Table S1.

We performed mtDNA hybridisation capture[22], [34] and Y chromosome STR analyses using AmpFlSTR® Yfiler® PCR Amplification Kit (Applied Biosystems). Y chromosome haplotype analysis was carried out in GeneMapper® ID Software v3.2.1 (Applied Biosystems).

Bioinformatic analyses

The *PAPline* package with default settings[35] 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 Table S1. Y chromosome haplogroups were classified using the *Yleaf* v2.2 software[36]. Mitochondrial contamination ratio was estimated with *contamMix* (v1.0-11)[37].

The ratio of male X-chromosomal contamination was measured, with both *ANGSD* (version 0.939-10-g21ed01c, htslib 1.14-9-ge769401)[38] and *hapCon* (*hapROH* package version 0.60)[39]. The starting file for the latter 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*[40] and *KIN*[41], and where available, supplemented by IBD connections (for results, see Table S2). For PCA, admixture analysis, and *f*-statistics, the dataset was merged with the AADR database v54[42]. PCA was calculated using the *smartpca* software[43] with the 590k SNP panel[42], for admixture analysis we used the *ADMIXTURE* software[44] with the 1240k SNP panel[42]. For *f*-statistics, we used the *Admixtools* software package[17], [45]. 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 the closely related individuals and those samples that were not treated with UDG.

We used *hapROH*[46] 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). We also used different ROH segment ranges (4-20 cM, 4-8 cM, 8-12 cM, 12-20 cM) to monitor whether they show different temporal trends, as larger segments signal more recent inbreeding or population size reduction. The results are listed in Table S8.

Individuals with >250k SNPs on the 1240k panel were targeted for IBD analysis. Imputation was performed using *GLIMPSE* software[21], 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[47]. 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*). The script used to phase and impute the samples is available on our github repository: https://github.com/ArchGenIn/AHP_Gerber2024/tree/main. To identify the IBD segments, we used *ancIBD* (version 0.5)[20] 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 IBD results were visualized using *Cytoscape*[48]. We restricted connections down to 1x10 cM, due to the high number of false-positive hits at shorter segments resulting from technical difficulties (*e.g.* phasing

errors) and the sparse and asymmetric background database for imputation discussed further in the description of Fig. S25.

We also measured degree centrality (*k*), using sum IBD connections between 10 and 1400 cM to exclude close relatives, as we were interested in the population-level processes. The conventional approach used for different time section batches (*i.e.* centuries) is the following for k_W:

$$\frac{1}{G}\sum_{i=1}^{G}\frac{2\cdot connections_i}{n_i\cdot(n_i-1)}$$

Where G is the number of groups within a batch, $connections_i$ is the number of IBD connections within the ith group (site) within the batch, and n is the number of individuals within the ith group.

For k_B the following formula was used:

$$\frac{1}{G(G-1)}\sum_{i=1}^{G}\sum_{j=1,i\neq j}^{G}\frac{connections_{ij}}{n_i\cdot n_j}$$

Where $connections_{ij}$ is the number of IBD connections between the pair of the ith and jth sites. For the cumulative approach (applied for k_B only), we used:

$$\frac{\sum_{i=1}^{G}\sum_{x=1,x\neq i}^{G} connections_{ix}}{n_i\cdot n_x}$$

Where $connections_{ix}$ denotes all IBD connections between the ith and all other sites within the batch.

We also used an alternate weighting method to highlight global trends visualized in Fig. S23:

$$\frac{\sum_{i=1}^{G}\sum_{x=1,x\neq i}^{G} connections_{ix}}{2n}$$

This latter normalization method balances sample size differences between sites and chronological batches, and thus highlights relative differences. The sample number distribution is summarized in Fig. S20. The script used for this calculation is also uploaded to our github repository.

References

[1] C. Bálint, *The Avars, Byzantium and Italy. A study in chorology and cultural history*, vol. 31. in Varia archaeologica Hungarica, vol. 31. Budapest, Hungary: Archaeolingua, 2019.

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

[3] Bóna I., Cseh J., Nagy M., Tomka P., and Tóth Á., *Hunok - Gepidák - Langobardok*. Szeged, Hungary: Magyar Őstörténeti Könyvtár, 1993.

[4] W. Pohl, *The Avars: A Steppe Empire in Central Europe, 567–822*. 2018. [Online]. Available: https://www.jstor.org/stable/10.7591/j.ctt21h4xd7

[5] B. M. Szőke, *DIE KAROLINGERZEIT IN PANNONIEN*. Mainz: Verlag des Römisch-Germanischen Zentralmuseums, 2021.

[6] T. Vida, "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, 2008.

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

[8] C. E. G. Amorim *et al.*, "Understanding 6th-century barbarian social organization and migration through paleogenomics," *Nat. Commun.*, vol. 9, no. 1, p. 3547, Dec. 2018, doi: 10.1038/s41467-018-06024-4.

[9] G. A. Gnecchi-Ruscone *et al.*, "Ancient genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites," *Cell*, vol. 185, no. 8, pp. 1402-1413.e21, Apr. 2022, doi: 10.1016/j.cell.2022.03.007.

[10] Z. Maróti *et al.*, "The genetic origin of Huns, Avars, and conquering Hungarians," *Curr. Biol.*, vol. 32, no. 13, pp. 2858-2870.e7, Jul. 2022, doi: 10.1016/j.cub.2022.04.093.

[11] D. N. Vyas *et al.*, "Fine-scale sampling uncovers the complexity of migrations in 5th–6th century Pannonia," *Curr. Biol.*, vol. 33, no. 18, pp. 3951-3961.e11, Sep. 2023, doi: 10.1016/j.cub.2023.07.063.

[12] V. Csáky *et al.*, "Genetic insights into the social organisation of the Avar period elite in the 7th century AD Carpathian Basin," *Sci. Rep.*, vol. 10, no. 1, p. 948, Dec. 2020, doi: 10.1038/s41598-019-57378-8.

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

[14] I. Mathieson *et al.*, "Genome-wide patterns of selection in 230 ancient Eurasians," *Nature*, vol. 528, no. 7583, pp. 499–503, Dec. 2015, doi: 10.1038/nature16152.

[15] V. Csáky *et al.*, "Early medieval genetic data from Ural region evaluated in the light of archaeological evidence of ancient Hungarians," *Sci. Rep.*, vol. 10, no. 1, p. 19137, Dec. 2020, doi: 10.1038/s41598-020-75910-z.

[16] M. L. Antonio *et al.*, "Stable population structure in Europe since the Iron Age, despite high mobility," *eLife*, vol. 13, p. e79714, Jan. 2024, doi: 10.7554/eLife.79714.

[17] N. Patterson *et al.*, "Ancient Admixture in Human History," *Genetics*, vol. 192, no. 3, pp. 1065–1093, Nov. 2012, doi: 10.1534/genetics.112.145037.

[18] I. Olalde *et al.*, "A genetic history of the Balkans from Roman frontier to Slavic migrations," *Cell*, vol. 186, no. 25, pp. 5472-5485.e9, Dec. 2023, doi: 10.1016/j.cell.2023.10.018.

[19] E. Yüncü *et al.*, "False discovery rates of qpAdm-based screens for genetic admixture," Genetics, preprint, Apr. 2023. doi: 10.1101/2023.04.25.538339.

[20] H. Ringbauer, Y. Huang, A. Akbari, S. Mallick, N. Patterson, and D. Reich, “ancIBD - Screening for identity by descent segments in human ancient DNA,” Genetics, preprint, Mar. 2023. doi: 10.1101/2023.03.08.531671.
[21] S. Rubinacci, D. M. Ribeiro, R. J. Hofmeister, and O. Delaneau, “Efficient phasing and imputation of low-coverage sequencing data using large reference panels,” *Nat. Genet.*, vol. 53, no. 1, pp. 120–126, Jan. 2021, doi: 10.1038/s41588-020-00756-0.
[22] B. Szeifert *et al.*, “Tracing genetic connections of ancient Hungarians to the 6th–14th century populations of the Volga-Ural region,” *Hum. Mol. Genet.*, vol. 31, no. 19, pp. 3266–3280, Sep. 2022, doi: 10.1093/hmg/ddac106.
[23] G. A. Gnecchi-Ruscone *et al.*, “Network of large pedigrees reveals social practices of Avar communities,” *Nature*, Apr. 2024, doi: 10.1038/s41586-024-07312-4.
[24] Juhász P., “A honfoglalók létszámának kérdéséhez,” *Belvedere Meridionale*, vol. 27, no. 3, pp. 66–76, 2015, doi: 10.14232/belv.2015.3.7.
[25] P. Langó, *Turulok és Árpádok: Nemzeti emlékezet és koratörténeti emlékek*. Budapest, Hungary: MTA BTK, 2017.
[26] Révész L., *A Karosi Honfoglalás Kori Temetők - Régészeti adatok a Felső-Tisza-vidék X. századi történetéhez*. Miskolc, Hungary, 1996.
[27] S. S. Ebenesersdóttir, Á. Sigurðsson, F. Sánchez-Quinto, C. Lalueza-Fox, K. Stefánsson, and A. Helgason, “A new subclade of mtDNA haplogroup C1 found in icelanders: Evidence of pre-columbian contact?,” *Am. J. Phys. Anthropol.*, vol. 144, no. 1, pp. 92–99, Jan. 2011, doi: 10.1002/ajpa.21419.
[28] Kovács G., *Középkori és Török kori várak, erődítések Pest megyében*. Budapest, Hungary: Archaeolingua Foundation, 2022.
[29] M. Lipson *et al.*, “Parallel palaeogenomic transects reveal complex genetic history of early European farmers,” *Nature*, vol. 551, no. 7680, pp. 368–372, Nov. 2017, doi: 10.1038/nature24476.
[30] K. A. Sirak *et al.*, “A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis,” *BioTechniques*, vol. 62, no. 6, pp. 283–289, Jun. 2017, doi: 10.2144/000114558.
[31] J. Dabney *et al.*, “Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 39, pp. 15758–15763, Sep. 2013, doi: 10.1073/pnas.1314445110.
[32] N. Rohland, I. Glocke, A. Aximu-Petri, and M. Meyer, “Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing,” *Nat. Protoc.*, vol. 13, no. 11, pp. 2447–2461, Nov. 2018, doi: 10.1038/s41596-018-0050-5.
[33] N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, and D. Reich, “Partial uracil–DNA–glycosylase treatment for screening of ancient DNA,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 370, no. 1660, p. 20130624, Jan. 2015, doi: 10.1098/rstb.2013.0624.
[34] N. Rohland, S. Mallick, M. Mah, R. Maier, N. Patterson, and D. Reich, “Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs,” *Genome Res.*, vol. 32, no. 11–12, pp. 2068–2078, Nov. 2022, doi: 10.1101/gr.276728.122.
[35] D. Gerber *et al.*, “Interdisciplinary Analyses of Bronze Age Communities from Western Hungary Reveal Complex Population Histories,” *Mol. Biol. Evol.*, vol. 40, no. 9, Aug. 2023, doi: https://doi.org/10.1093/molbev/msad182.
[36] A. Ralf, D. Montiel González, K. Zhong, and M. Kayser, “Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data,” *Mol. Biol. Evol.*, vol. 35, no. 5, pp. 1291–1294, May 2018, doi: 10.1093/molbev/msy032.

[37] Q. Fu *et al.*, “A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes,” *Curr. Biol.*, vol. 23, no. 7, pp. 553–559, Apr. 2013, doi: 10.1016/j.cub.2013.02.044.
[38] T. S. Korneliussen, A. Albrechtsen, and R. Nielsen, “ANGSD: Analysis of Next Generation Sequencing Data,” *BMC Bioinformatics*, vol. 15, no. 1, p. 356, Dec. 2014, doi: 10.1186/s12859-014-0356-4.
[39] Y. Huang and H. Ringbauer, “hapCon: estimating contamination of ancient genomes by copying from reference haplotypes,” *Bioinformatics*, vol. 38, no. 15, pp. 3768–3777, Aug. 2022, doi: 10.1093/bioinformatics/btac390.
[40] J. M. Monroy Kuhn, M. Jakobsson, and T. Günther, “Estimating genetic kin relationships in prehistoric populations,” *PLOS ONE*, vol. 13, no. 4, p. e0195491, Apr. 2018, doi: 10.1371/journal.pone.0195491.
[41] D. Popli, S. Peyrégne, and B. M. Peter, “KIN: a method to infer relatedness from low-coverage ancient DNA,” *Genome Biol.*, vol. 24, no. 1, p. 10, Jan. 2023, doi: 10.1186/s13059-023-02847-7.
[42] S. Mallick and D. Reich, “The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes,” vol. Harvard Dataverse, no. 8, 2023, doi: https://doi.org/10.7910/DVN/FFIDCW.
[43] N. Patterson *et al.*, *EIGENSOFT*. (2017). [Linux]. Available: https://github.com/DReichLab/EIG
[44] D. H. Alexander, J. Novembre, and K. Lange, “Fast model-based estimation of ancestry in unrelated individuals,” *Genome Res.*, vol. 19, no. 9, pp. 1655–1664, Sep. 2009, doi: 10.1101/gr.094052.109.
[45] R. Maier, N. Patterson, and D. Reich, *admixtools*. (2021). R. [Online]. Available: https://github.com/uqrmaie1/admixtools
[46] H. Ringbauer, J. Novembre, and M. Steinrücken, “Parental relatedness through time revealed by runs of homozygosity in ancient DNA,” *Nat. Commun.*, vol. 12, no. 1, p. 5425, Sep. 2021, doi: 10.1038/s41467-021-25289-w.
[47] H. Li *et al.*, “The Sequence Alignment/Map format and SAMtools,” *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, Aug. 2009, doi: 10.1093/bioinformatics/btp352.
[48] P. Shannon *et al.*, “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks,” *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003, doi: 10.1101/gr.1239303.
[49] M. Héder *et al.*, “The Past, Present and Future of the ELKH Cloud,” *Inf. Társad.*, vol. 22, no. 2, p. 128, Aug. 2022, doi: 10.22503/inftars.XXII.2022.2.8.
[50] Horváth C., *HONFOGLALÁS ÉS KORA ÁRPÁD-KORI SÍROK, TEMETŐK ÉS SZÓRVÁNYLELETEK A NYUGAT-DUNÁNTÚLON I.* in Magyarságkutató Intézet Kiadványai, no. 52. Budapest, Hungary, 2022.
[51] P. Tomka, “Teil eines Gräberfeldes aus der Karolingerzeit von Himod, Flur Káposztás.,” *Anteus*, pp. 199–223, 2010.
[52] K. Éry, “Reconstruction of the tenth century population of Sárbogárd on the basis of archeological and anthropological data. – A sárbogárdi X. századi közösség rekonstrukciója régészeti és embertani adatok alapján.,” *Alba Regia*, vol. 8–9, pp. 93–147, 1967.
[53] A. Kralovánszky, “DIE LANDNAHMEZEITLICHE RINDERBESTATTUNG VON SÄRBOGÄRD,” *Alba Regia*, no. 6–7, pp. 89–99, 1966.
[54] A. Kralovánszky, “Szarvasmarha temetkezés a honfoglalás korából,” *Magy. Mezőgazdasági Múz. Közleményei I*, pp. 171–184, 1964.

[55] P. Gróf and I. Kováts, “Árpád-kori temetőrészlet és XIV. századi hypocaustum Visegrád belterületéről – An Árpádian age cemetery and a 14th century hypocaust at Visegrád,” *Commun. Archaeol. Hung.*, pp. 183–196, 2014.
[56] J. Kodolányi, “Templomok és temetők a visegrádi Sibrik-dombon,” in *Genius loci. Laszlovszky 60*, ARCHAEOLINGUA ALAPÍTVÁNY, 2018, pp. 181–185.
[57] G. Buzás, K. Boruzs, S. Merva, and K. Tolnai, “Régészeti kutatások a visegrádi Sibrik-dombon. - Archaeological investigations on Sibrik Hill at Visegrád.,” *Commun. Archaeol. Hung.*, pp. 193–235, 2017.
[58] E. Kiss, “Eirene and the Holy Theotokos,” in *Piroska and the Pantokrator*, CEU MEDIEVALIA, 2019, pp. 268–269.
[59] Mecsi B., “Magyarország egyik legrégibb festészeti emléke: A visegrádi esperesi templom falképtöredékei,” in *Ars Perennis*, Budapest, Hungary, 2010, pp. 19–21.
[60] K. Bakay, “Gräberfelder aus den 10–11. Jahrhunderten in der umgebung von Székesfehérvár und die Frage der Fürstlichen Rezidenz,” *Alba Regia*, no. 6–7, pp. 43–88, 1966.
[61] A. Marosi, “A székesfehérvári rádiótelepi ásatás,” *Archeol. Ért.*, vol. Új folyam, no. 40, pp. 245–257, 2926 1923.
[62] A. Marosi, “Székesfehérvár honfoglaláskori temetői,” *Archeol. Ért.*, vol. Új folyam, no. 39, pp. 25–41, 1922 1920.
[63] J. Fitz, *A Fejér Megyei Múzeumok évi jelentése 1956. évre*. in István Király Múzeum Közleményei, no. E/2.5. 1957.
[64] Z. Bánki, “Régészeti Kutatások,” *Alba Regia*, vol. 12, pp. 280–285, 1972.
[65] Z. Bánki, “Régészeti Kutatások,” *Alba Regia*, vol. 11, pp. 163–166, 1971.
[66] A. Kralovánszky, “Székesfehérvár-Ikarus Gyár,” *Régészeti Füzetek*, vol. 47, no. I/22, 1969.
[67] Z. Petkes, “Kora Árpád-kori temetőrészlet Bodajk_Proletár-földek II. lelőhelyről,” in *Avarok és magyarok Bodajkon. 6-10. századi temetkezések komplex elemzése*, vol. 60, in A Szent István Király Múzeum Közleményei, no. A, vol. 60. , Székesfehérvár, Hungary: Szent István Király Múzeum, 2021, pp. 247–264.
[68] F. Szücsi, *Avarok és magyarok Bodajkon. 6-10. századi temetkezések komplex elemzése*, vol. 60. in A Szent István Király Múzeum Közleményei, no. A, vol. 60. Székesfehérvár, Hungary: Szent István Király Múzeum, 2021.
[69] A. Marosi and N. Fettich, “Dunapentelei avar sírleletek - Trouvailles avares de Dunapentele,” *Archaeol. Hung.*, vol. 18, 1936.
[70] J. Fitz, “Hogyan folyik a régészeti ásatás Előszálláson?,” *Fejér Megyei Népl.*, vol. 8, 1953.
[71] É. Petres and J. Fitz, *Fejér megye rövid története I. - A legrégebbi időktől a honfoglalás koráig*. Székesfehérvár, Hungary: István Király Múzeum Közleményei, 1956.
[72] T. Vida, *Die Awarenzeitliche Keramik I*. in Varia archaeologica Hungarica. Berlin-Budapest: Balassi, 1999.
[73] J. Fitz, “Előszállás környékének régészeti topográfiája,” *Székesfehérvár Szle.*, vol. 10, pp. 13–21, 1955.
[74] S. Wenger, “Anthropologie de la population d’Előszállás-Bajcsihegy. Provenant des temps Avers,” *Anthropol. Hung.*, vol. 7, no. 1–2, pp. 115–164, 1966.
[75] G. László, *Das awarenzeitliche Gräberfeld in Csákberény-Orondpuszta*. in Monumenta Avarorum Archaeologica. Magyar Nemzeti Múzeum, 2015.

[76] F. Szücsi, "„Az ismeretlen ismerős": Új eredmények a csákberény-orondpusztai avar kori temető kutatásából: a sírépítmények," *Hadak Útján*, pp. 369–397, 2022, doi: 10.55722/Arpad.Kiad.2021.4.1_19.
[77] F. Szücsi, "Csákberény, Orondpuszta, tervásatása," *Alba Regia*, vol. 48, pp. 197–198, 2020.
[78] A. Kralovánszky, "A Pákozd-Börgöndpusztai későavarkori temető - Spätawarenzeitliches Gräberfeld in Pákozd-Börgöndpuszta," *Alba Regia*, vol. 2–3, pp. 173–179, 1963.
[79] E. Neparáczki *et al.*, "Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing," *PLOS ONE*, vol. 12, no. 4, p. e0174886, Apr. 2017, doi: 10.1371/journal.pone.0174886.
[80] C. Jeong *et al.*, "A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe," *Cell*, vol. 183, no. 4, pp. 890-904.e29, Nov. 2020, doi: 10.1016/j.cell.2020.10.015.
[81] I. Lazaridis *et al.*, "Ancient human genomes suggest three ancestral populations for present-day Europeans," *Nature*, vol. 513, no. 7518, pp. 409–413, Sep. 2014, doi: 10.1038/nature13673.
[82] M. L. Antonio *et al.*, "Stable population structure in Europe since the Iron Age, despite high mobility," Genetics, preprint, May 2022. doi: 10.1101/2022.05.15.491973.

Acknowledgement

Dr. Mihály Molnár (MTA ATOMKI HEKAL, AMS laboratory) is responsible for the radiocarbon analyses from Visegrád site.

On behalf of the 'bioinfo-baden' project we are grateful for the computational possibility to the HUN-REN Cloud[49] (https://science-cloud.hu/), which helped us achieve the results published in this paper.

We'd like to thank Ali Akbari for his help with the imputation of the whole genome data. We also thank Harald Ringbauer and Yilei Huang for their suggestions on running and optimizing ancIBD.

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-Vársziget), as well as anthropologist Tamás Szeniczey. We would like to thank for the invaluable comments on the paper to Péter Langó, Tivadar Vida, István Koncz, Balázs Gyuris and Levente Samu.

Funding

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).

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.

The radiocarbon investigations were carried out in the framework of the OTKA NK 104533 research project of Miklós Takács, which we are grateful for.

Author contribution

Ethics Statement

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

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.

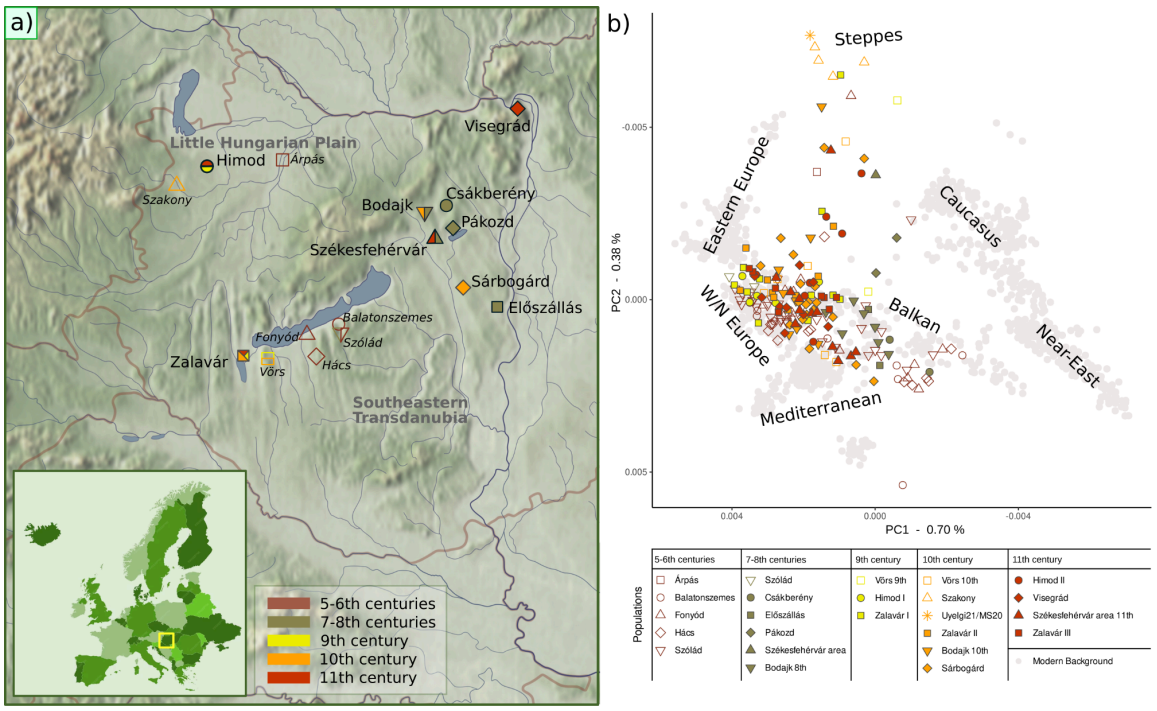

Figure 1. Sites and general genomic composition of the samples originating from Transdanubia. *a) The discussed sites in Transdanubia, where italic labels and empty symbols indicate previously published datasets[8], [10], [11]. b) Principal component*

analysis (PCA) presenting the Transdanubian genomes from the 5th to 11th centuries CE. Ancient published (empty symbols) and newly sequenced (filled symbols) genomes are projected onto modern West Eurasian genomes (grey dots).

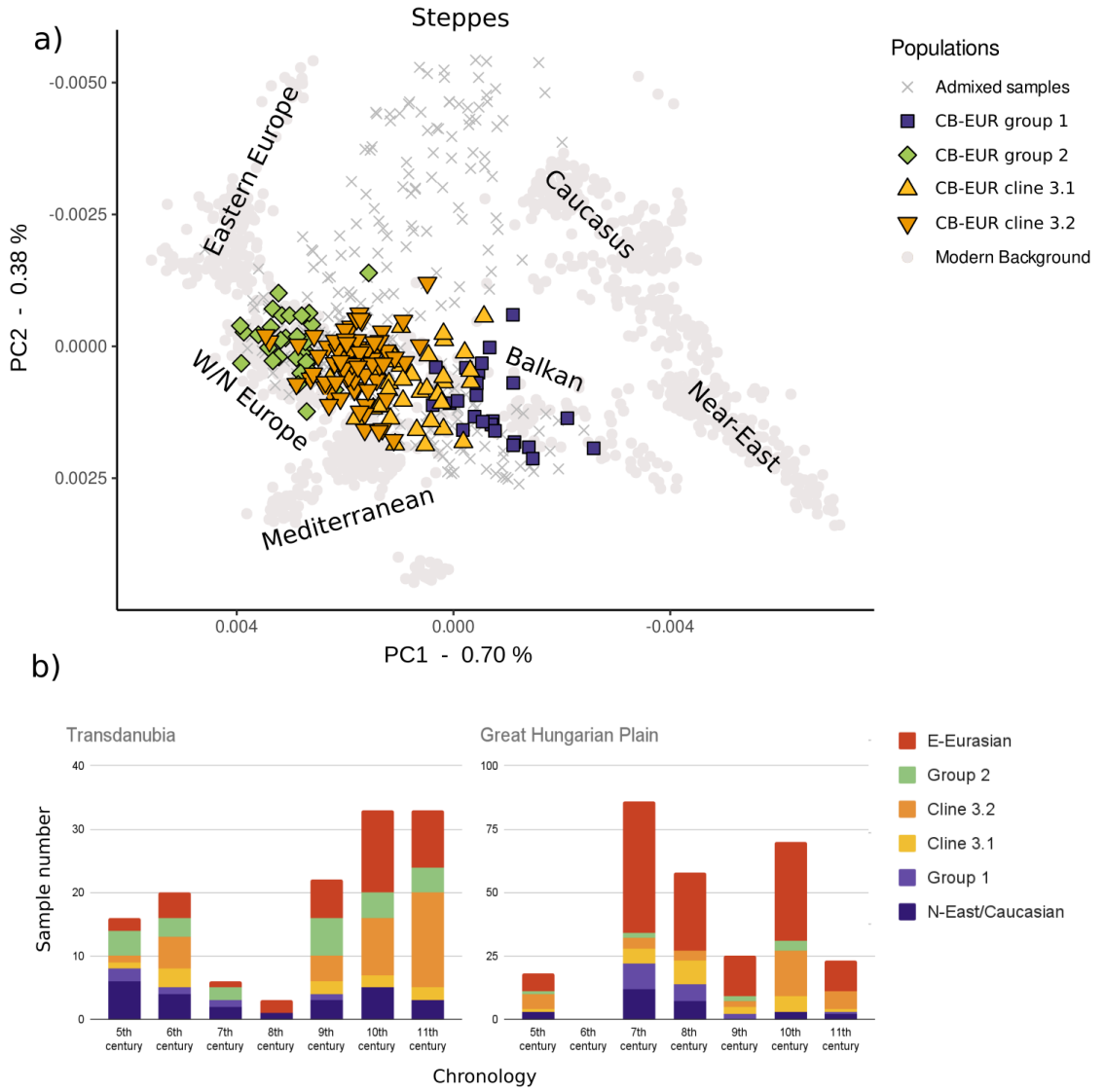

Figure 2. Distribution of European genetic groups. *a) Principal component analysis (PCA) representation of f_4-cluster based classification of individuals bearing only European ancestry. Further tests (Supplementary Information section 3.2) revealed that these groups represent sections of a regional 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.*

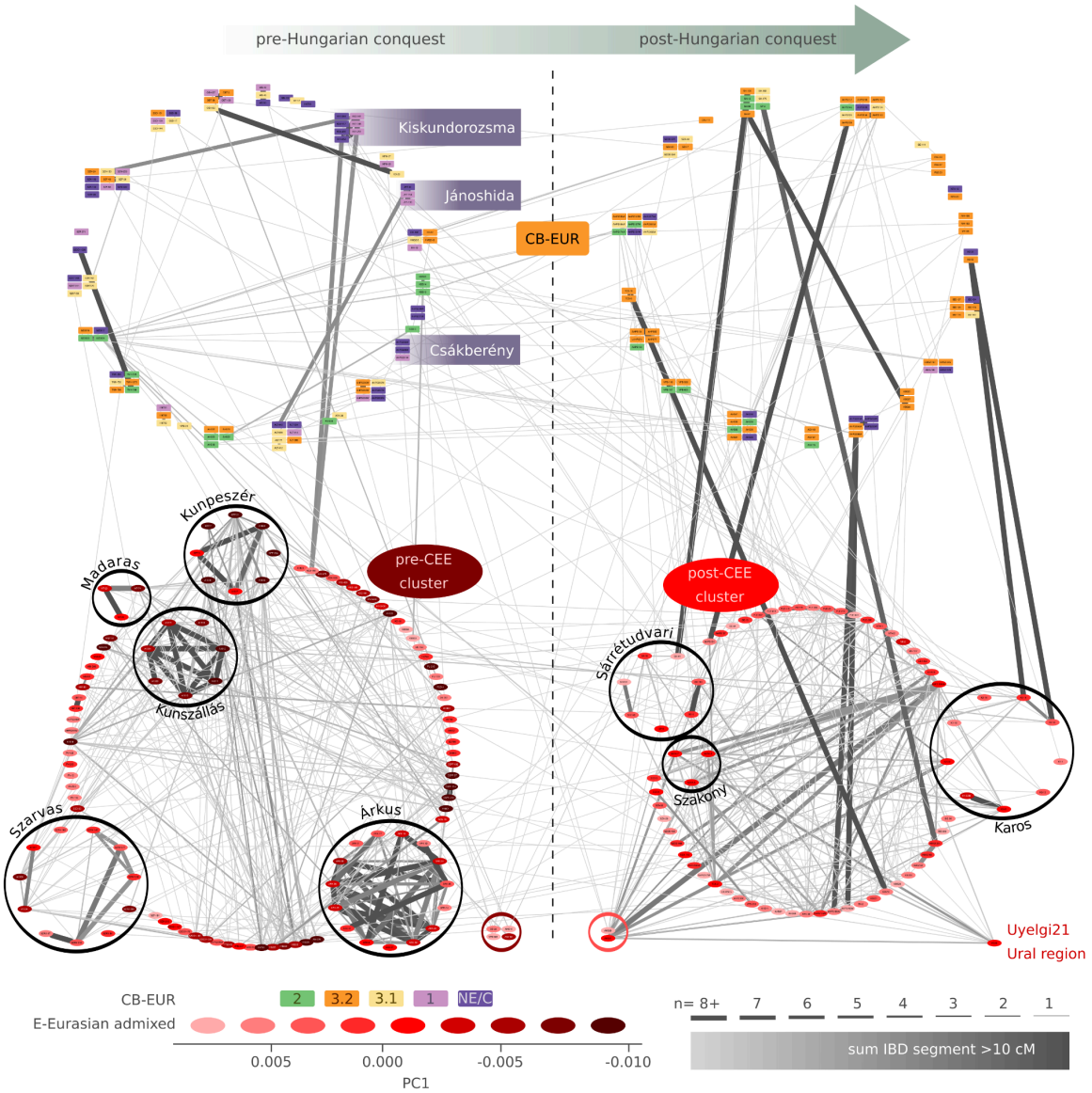

Figure 3. IBD (identity-by-descent) segment sharing networks of the Carpathian Basin populations in the 7th-11th centuries CE. *The colouration of the individuals in CB-EUR corresponds to their European genetic groups/clines. The relative positions of the first principal component (PC1) are illustrated at the bottom. The samples are grouped by sites (Table S5). Ovals in the lower clusters represent individuals with significant East-Eurasian ancestry and color depth reflects their PC1 positions. Samples are sorted by sites, with those mentioned in the text circled in black. The width of the edges corresponds to the number of >10 cM IBD segments and the color depth signals the total shared length in cM (centimorgan). Three sites are highlighted for 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, refer to Table S6.*

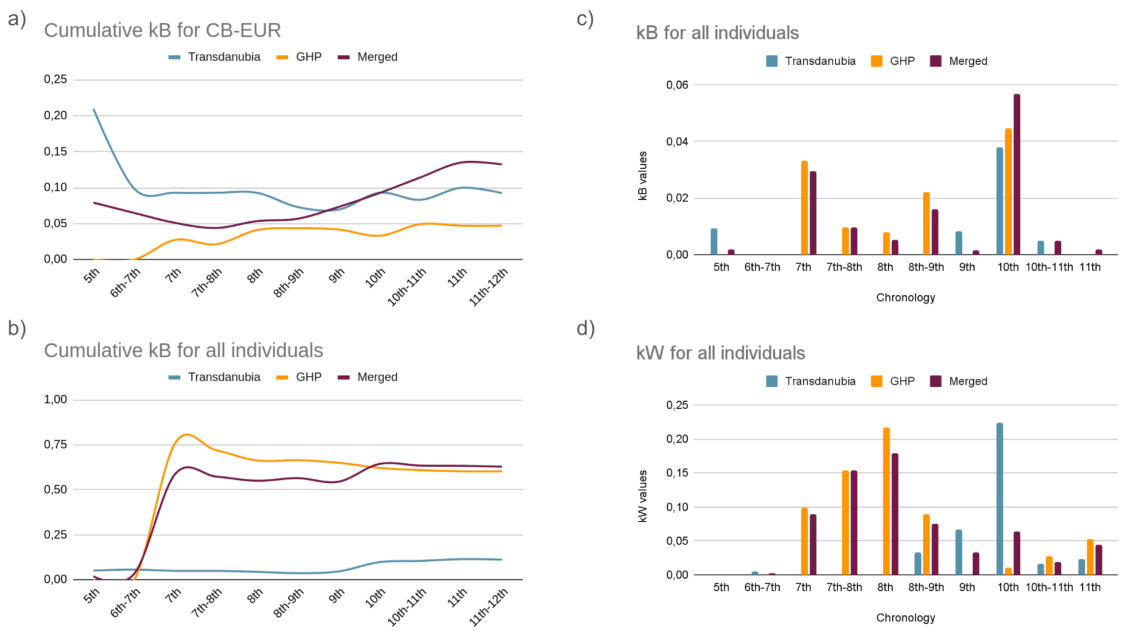

Figure 4. Intra- and intergroup tendencies in IBD (identity-by-descent) segment sharing patterns. *a) Cumulative connections between sites (k_B) regarding the CB-EUR cluster in a chronological order, indicate movement into and within Transdanubia during the 10th century CE but not within the GHP (Great Hungarian Plain). b) The same analysis, when including all individuals (also CEE), signals contact between sites in the GHP during 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. d) Within site connections (k_W) reveal moderate biological networks across sites (Table S6.2).*

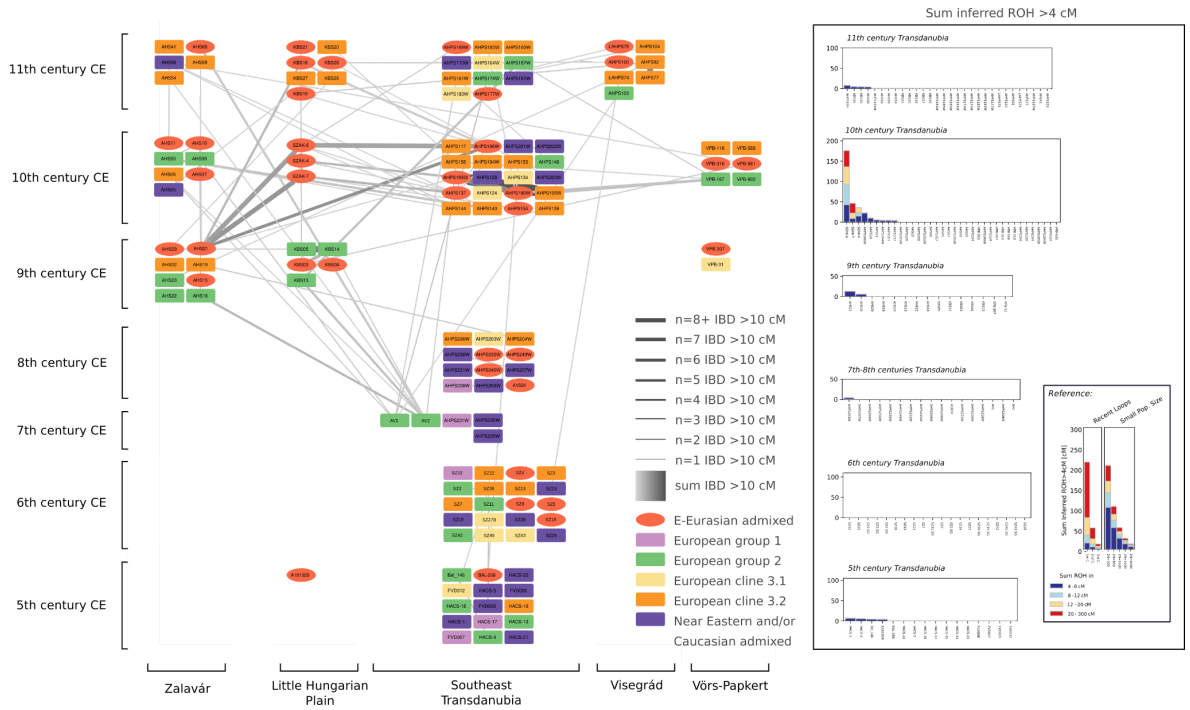

Figure 5. IBD (identity-by-descent) patterns across Transdanubian sites in the 5th-11th centuries and their corresponding ROH patterns. *Only connections between 10 and 1400 cM (centimorgan) are visualized, in order to emphasize population level connections instead of 1-2nd degree connections.*

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.[15]