PhD DISSERTATION

The origin and composition of the "Forgotten people": genetic analysis of the Sarmatian-period population of the Carpathian Basin

By: Oszkár Schütz

Supervisor:

Tibor Török, PhD
ASSOCIATE PROFESSOR

DOCTORAL SCHOOL OF BIOLOGY

DEPARTMENT OF GENETICS
FACULTY OF SCIENCE AND INFORMATICS
UNIVERSITY OF SZEGED

SZEGED
2025

TABLE OF CONTENT

I INTRODUCTION

I.1 History and archaeology of the Sarmatians

I.1.1 The Steppe connections and migration of the Sarmatians

By the middle of the 1st century CE, the political and ethnic landscape of the Carpathian Basin underwent significant changes. By the end of the previous century, the dominant political entity in the region had become the Dacian Kingdom, centered in Transylvania, which united primarily Dacian and Getae tribes under its rule. In addition to Dacian expansion into the territories of the late Iron Age Celtic and Scythian populations living in the eastern part of the Carpathian Basin, the new century saw the emergence of political formations that would shape the region's history for centuries. Roman administration gradually established its system in Transdanubia (the province of Pannonia), while north of it, the first Germanic-speaking tribes (the Quadi and the Marcomanni) appeared. Meanwhile, taking advantage of the temporary weakening of the Dacian state, the first Iranian-speaking groups, the Sarmatians, moved into the northern strip of the Duna-Tisza Interfluve (Bârcă & Symonenko, 2009; Istvánovits & Kulcsár, 2006).

The Sarmatian tribes (the Aorsi, Siraci, Roxolani, and Iazyges) had already been influential groups in the Eastern European steppe since the 3rd-2nd centuries BCE. Culturally, in terms of their lifestyle and social structure, they were closely tied to the Scythian groups that preceded them (Melyukova, 1990; Mordvintseva, 2013; Tokhtasyev, 2005). The Sarmatian tribes, who migrated from the Ural and Volga regions north of the Black Sea, took over the role of the Scythians within the nomadic world after the collapse of Scythian power and brought the steppe under their control. Among the confederations of tribes collectively referred to by the ancient world as "Sarmatians," the Iazyges and Roxolani, living furthest to the west, played a prominent role in the history of the Carpathian Basin. According to written sources, the Iazyges, arriving from the lower Danube region, were the first to enter the Great Hungarian Plain, when they became mercenaries for Vannius, king of the Quadi, around 50 CE (Istvánovits & Kulcsár, 2006). These communities exploited a temporary political situation, as neither the Roman political and administrative system had yet fully established itself following Roman expansion, nor had Dacian power fully recovered from its brief decline.

The first archaeological materials associated with these groups date from the end of the 1st century to the beginning of the 2nd century, showing a slight temporal shift compared to the

written sources. The early Sarmatian communities are primarily represented by small family cemeteries and the graves of women and children, often containing gold ornaments brought from the east, forming what is referred to in the archaeological literature as the "gold horizon" of burials. Alongside these, male elite figures, associated with steppe artifacts such as tamga-marked gold fittings and ring-hilted swords, also emerge around the turn of the century (e.g., Dunaharaszti, Újszilvás) (H. Vaday, 1989b; Tari, 1994). The material culture of the first migrating generation, showing nomadic and Black Sea regional influences, was also impacted by the local Celtic and early Roman culture, as well as the regional Dacian culture (Istvánovits & Kulcsár, 2006; Masek, 2016). However, the early elite burials often contained objects reflecting international connections with more distant regions (e.g., Vistula region, Veresegyház) (Masek, 2016; Mesterházy, 1986). By the end of the 1st century, the Sarmatians played an increasingly important role in local political and diplomatic affairs, becoming regular participants in the anti-barbarian wars led by Emperor Domitian (Istvánovits & Kulcsár, 2017).

I.1.2 The Sarmatians' relations with neighbouring peoples

After the fall of the Dacian Kingdom following Roman military campaigns in 106 CE, Sarmatian settlements, as indicated by archaeological findings, extended eastward beyond the Tisza River and began to populate the southern areas of the Great Hungarian Plain near the Roman frontier (Bácska and Banat, now northern Serbia) (Grumeza, 2014; Istvánovits & Kulcsár, 2017). Even at the time of their first appearance, the Sarmatian communities in the Great Plain were part of a broader international network, as evidenced by their earliest burial finds (e.g., Dabas, Veresegyház). During the early Sarmatian Period, while Eastern-style material culture and burial traditions were prevalent, the greatest influences came from Dacian and Roman provincial cultures.

The populations migrating from the steppe into the Carpathian Basin quickly adapted to local conditions, abandoning their nomadic lifestyle relatively swiftly. This transition is evidenced by the extensive settlement network that developed and the archaeological findings indicating significant agricultural and artisanal activities. (Istvánovits et al., 2005). Although several elements of their burial customs remained tied to steppe traditions (e.g., graves with ditch enclosures, burial mounds, and incense burners), these practices saw little change over the centuries (Kulcsár, 1998b). The material culture of the Sarmatians in the Great Plain incorporated many local influences, especially in clothing, weapons, and pottery, which bore marks of Celtic, Dacian, and Roman impact (Istvánovits & Kulcsár, 2020).

In addition to elements of early Celtic-Dacian pottery traditions, the Sarmatians increasingly came under Roman influence from the 2nd century onward. As they expanded, they likely absorbed local populations throughout the century (H. Vaday, 1991, 1996). During the Marcomannic Wars (166–180 CE), the Sarmatian groups in the Great Plain launched intense attacks on neighboring Roman provinces (Pannonia, Dacia, Moesia). Throughout the war and its aftermath, the movement of small warrior bands outside the Roman Empire likely intensified, as they sought plunder near the empire's borders. Despite the Roman victory and the harsh peace terms (e.g., trade restrictions and mandatory conscription), by the end of the 2nd century, the Sarmatians emerged as one of the most significant groups in the region. Unlike the Germanic tribes, their territories were not occupied by the Romans during the conflicts, which contributed to their continued prominence (P. Kovács, 2009).

The settlement of a significant portion of the Great Hungarian Plain was completed during this period, and a network of interconnected settlements was established. From the late 2nd to the mid-3rd century, their relations with the Romans and Germans flourished, as evidenced by the influx of imported goods into the Great Hungarian Plain (Bene et al., 2016, p. 20; H. Vaday, 1989b; H. Vaday & Horváth, 2005; Istvánovits & Kulcsár, 2020). Roman items appearing in Sarmatian territories were quickly replicated and adapted to their local culture. The 2nd and 3rd-century local material culture can be interpreted as a kind of peripheral Roman culture in the Barbaricum. These close ties with Rome laid the economic foundation for producing Roman-influenced goods locally during the 3rd and 4th centuries.

It is also likely that new Sarmatian groups migrated from the east to the Great Hungarian Plain during and after the Marcomannic Wars, as seen in the distinct archaeological finds and new Eastern burial rites (e.g., the Hévízgyörk-Vizesdpuszta group) (Dinnyés, 1991; Istvánovits & Kulcsár, 2017; Khrapunov, 2001; Kulcsár, 1998a). Alongside increasing Roman contacts, a mixed and blended culture also emerged from the 2nd century onward in regions neighboring the Germanic peoples (Istvánovits & Kulcsár, 2000, p. 200, 2003; Vaday, 2001). In regions adjacent to the Quadi and Przeworsk culture territories, objects typical of Germanic culture were found, with similar items increasingly appearing further south in the Great Hungarian Plain. This can be seen as the blending of cultures (as reflected in recent finds and scientific data from the Hódmezővásárhely region) (S. Varga, 2020).

During this period, a new neighbor appeared near the Upper Tisza region—the Przeworsk culture (the archaeological culture of the Vandals)—with whom the Sarmatians developed intensive relations in their northern periphery (Soós, 2019a). By the late 3rd century, pressure

on the Sarmatian groups increased due to the migration of Gothic tribes into the Black Sea region. Internal tensions manifested in increased raids on Roman territories and internal conflicts. The Roman state also intervened in the affairs of the late Sarmatian Period in the Great Plain, building defensive structures such as parts of the so-called Csörsz Ditch.

The 4th century saw further population growth, likely due to the arrival of new groups from the east. Alongside strong late Roman connections, the material culture of the Great Hungarian Plain was also heavily influenced by the Marosszentanna-Chernyakhov culture (an archaeological culture that emerged in Eastern Europe from the 3rd century, incorporating both earlier Sarmatian populations and newly settled Eastern Germanic groups). These influences became more prominent in the archaeological record of the Great Hungarian Plain during this period (Bierbrauer, 1994, 1999; Körösfői, 2024).

I.1.3 The Sarmatians' lifestyle in the Great Hungarian Plain

The Sarmatian populations, who appeared in the Carpathian Basin during the 1st and 2nd centuries CE, brought with them a nomadic way of life from the steppe regions above the Black Sea and the Lower Danube. However, the newly acquired territory was not conducive to sustaining their previous lifestyle for long periods, forcing the Sarmatian groups to gradually adopt a more settled and complex way of life. Archaeological evidence of larger, permanent settlements begins to appear by the mid-2nd century, but it was during the 3rd and 4th centuries that these large, village-like settlements reached their peak across the Great Hungarian Plain (Istvánovits et al., 2013).

Despite the shift to a settled lifestyle with the emergence of agriculture, large-scale animal husbandry continued to play a significant role in the everyday livelihood of the Sarmatians in the Carpathian Basin. Based on current botanical and archaeozoological data, the cultural foundations of steppe nomadism persisted throughout the period (Istvánovits & Kulcsár, 2015). Archaeological findings primarily reflect the cultivation of crops typical of nomadic agriculture (barley, emmer wheat, millet), and the livestock, dominated by cattle and small ruminants like sheep, also exhibits characteristics typical of nomadic practices.

Due to their trade and cultural interactions with the Romans, certain crops and animal breeds from the provinces also appeared in regions near the frontier. Additionally, crossbreeding between Roman and local livestock is evident in some of these areas (Istvánovits & Kulcsár, 2015; Masek, 2021; Pető et al., 2017). Nevertheless, the Sarmatians maintained strong elements

of their nomadic heritage, particularly in agricultural and pastoral practices, even as they adapted to a more settled way of life on the Great Hungarian Plain.

I.1.4 Changes in the 4th–5th centuries and the possible continuity of the Sarmatians

From the latter half of the 4th century onwards, the Barbaricum (non-Roman territories) of the eastern Carpathian Basin underwent significant transformations. The increasing pressure of the Huns' advance, beginning in the 370s, triggered a large-scale migration across the former imperial regions of Central Europe. Tensions had already been building before the Huns' arrival, evident from attacks on Roman borders and changes in archaeological finds. The Hunnic migration resulted in cultural shifts, with earlier archaeological cultures showing foreign influences in their final phases, likely due to the movement of smaller groups (Pinar & Jiřík, 2019).

In the Great Hungarian Plain, these changes were reflected in the appearance of new objects and customs in Sarmatian archaeological sites, indicating the presence of new populations. During this period, more weapon burials were discovered compared to previous centuries. The influence of the Marosszentanna-Chernyakhov culture (from the east) and the Przeworsk culture (associated with Vandals and other Germanic groups) also became evident. Previously, research had focused on northeastern Hungary (e.g., the Tiszadob group) to track these changes, but similar phenomena have now been found throughout the Great Hungarian Plain (Istvánovits & Kulcsár, 2000), reflecting the blending of different cultures and preparing the way for the unified Hunnic period (Soós, 2019b; Tejral, 2011).

By the early 5th century, written sources and archaeological evidence suggest that part of the earlier Roman-period population left the Carpathian Basin (e.g., Radagaisus' invasion, the westward migration of the Alans, Vandals, and Suebi). The late Sarmatian culture is primarily identified through settlements, which continued to be occupied into the Hunnic period (H. Vaday, 1994; Masek, 2021; Tejral, 2011). Some of the larger cemeteries may have been abandoned around the early 5th century, but many settlements show continuity, even after the mid-5th century, despite new influences.

In the southern regions of the Great Hungarian Plain (around Szeged and Csongrád), archaeological sites from the Hunnic period display a combination of contemporary artifacts with Sarmatian traditions, especially in burial customs and material culture (Párducz, 1959, 1963). Written sources suggest that a significant portion of the Sarmatian population remained

in the area. Even after the collapse of the Hunnic Empire, records indicate the presence of Sarmatian leaders in the region. The last known reference to an independent Sarmatian political structure dates to the 470s. After this, it is likely that the remaining Sarmatian population assimilated into the emerging Gepid Kingdom (Kiss, 2015).

Thus, despite the dramatic upheavals of the 4th and 5th centuries, there is evidence of Sarmatian cultural continuity in parts of the Carpathian Basin, especially in their settlements, burial practices, and interactions with new groups like the Huns and Gepids.

I.1.5 Chronology of the Sarmatian Period in archaeological research

The Sarmatian Period has different chronological frameworks in Eastern Europe and the Carpathian Basin, which often do not align. In the vast region between the Ural Mountains and the Carpathian passes, Russian and Ukrainian archaeological research during the 20th century considered the peak of Sarmatian culture to span from the 2nd century BCE to the end of the 4th century CE.

Hungarian archaeological research first began focusing on Sarmatian groups in the early 20th century. The first attempts at distinguishing territorial and chronological groups were made by Mihály Párducz, who based his analysis on burial finds (Párducz, 1941, 1944, 1950). Párducz concentrated on the connections between the nomads of the eastern steppes and other cultures in the Carpathian Basin. He combined these perspectives with his observations on burial customs to establish regional and chronological groups, though these categorizations did not stand the test of time.

Later, in the 1980s and 1990s, more comprehensive syntheses emerged, which remain influential in Hungarian Sarmatian archaeology today. In 1989, Andrea Vaday created a chronological framework for the Sarmatian Period in the Carpathian Basin by carefully analyzing the material finds and drawing on the better-researched Roman and Germanic chronologies. (H. Vaday, 1989a). Vaday also incorporated data from historical sources to define broader chronological divisions, often tied to significant milestones in Sarmatian-Roman interactions. The Sarmatian chronological system of the Carpathian Basin have been further developed at the level of individual objects and phenomena in recent decades by Eszter Istvánovits and Valéria Kulcsár, who, alongside Eastern analogies, extensively utilized the Roman and Germanic artifacts that were available (Istvánovits & Kulcsár, 2017).

The Carpathian Basin's eastern half has its own distinct archaeological chronology, which poses several challenges from both Central-Northern European and steppe perspectives. For regions

outside the Roman Empire, such as the Barbaricum, the chronological framework established in the second half of the 20th century for Germanic territories was typically applied (Godłowski, 1970).

This however is not the case for the Carpathian Basin Sarmatian Period. Based on the main historical events of Roman-Sarmatian relations and changes in material culture, previous archaeological research has classified Sarmatian archaeological remains in the Carpathian Basin into three chronological periods. These periods are:

1. **Early Sarmatian Period**: From the arrival of the Sarmatians in the Great Hungarian Plain (second half of the 1st century CE) to the 2nd half of the 2nd century CE.
2. **Middle Sarmatian Period**: From the period of the Marcomannic wars to the end of the 3rd century CE abandonment of Dacia.
3. **Late Sarmatian Period**: From the end of the 3rd century CE to the last third of the 5th century CE.

To date, only one attempt (by Margit Nagy, in the case of the Pécel cemetery) has been made to integrate the archaeological development of Sarmatian groups in the Great Hungarian Plain into the wider chronological system of the Barbaricum (Nagy, 2018). This attempt was motivated by the strong Germanic influences at the site. Future research may likely follow this path, as the Sarmatians of the region were closely intertwined with the cultural and economic context of their Germanic neighbors.

Recently, Lavinia Grumeza has sought new approaches to partially synchronize the Hungarian archaeological chronology with the northern and central European Germanic material. Additionally, Norbert Kapcsos has applied mathematical methods to develop a new chronological framework for sites along the lower Maros (Mureș) River, contributing further to the study of the Sarmatian Period in the region (Grumeza, 2014; Kapcsos, 2022).

I.2 Previous genetic investigations concerning Sarmatians

I.2.1 The emergence of a unified Steppe ancestry

The Middle-Late Bronze Age period (~2200-1500 BC) on the Eurasian steppe saw an unprecedented wave of homogenisation with a genetic composition remarkably similar to that of the Western European populations form the same age (Narasimhan et al., 2019). This genetic composition is seemingly a stable mixture of distinct ancestral components representing three ancient metapopulations that progressively dominated the western end of the Eurasian continent from the last Glacial Maximum onward. This includes Western Hunter-Gatherers (WHG, green on **Figure 1**; Lazaridis et al., 2014), the original Mesolithic inhabitants of Western-Europe; European Neolithic Farmers (ANAT_N, yellow on **Figure 1**; Mathieson et al., 2018), the descendants of Neolithic migrants that came to dominate Europe after the Neolithic Expansion (~7000 BC); and descendants of the Yamnaya people (Yamnaya on **Figure 1**; Allentoft et al., 2015) who originated in the Pontic Steppe region and began their expansion into Europe at the onset of the Early Bronze Age (~ 3300 BC). The latter group also consists of two distinct genetic components: they derived a significant portion of their ancestry from the sparsely sampled Ancient North Eurasian metapopulation (ANE, red on **Figure 1** represented by hunter-gatherers from Russia; Narasimhan et al., 2019; Raghavan et al., 2014) that inhabited large territories of Northern and Central Eurasia; and Caucasus Hunter Gatherers (Jones et al., 2015) that exhibit very similar genetic composition to the Neolithic farmers of ancient Iran and Turan (IRAN_N, purple on **Figure 1**; for further discussion on the blue component see **III.2.2** and **IV.3.2**).

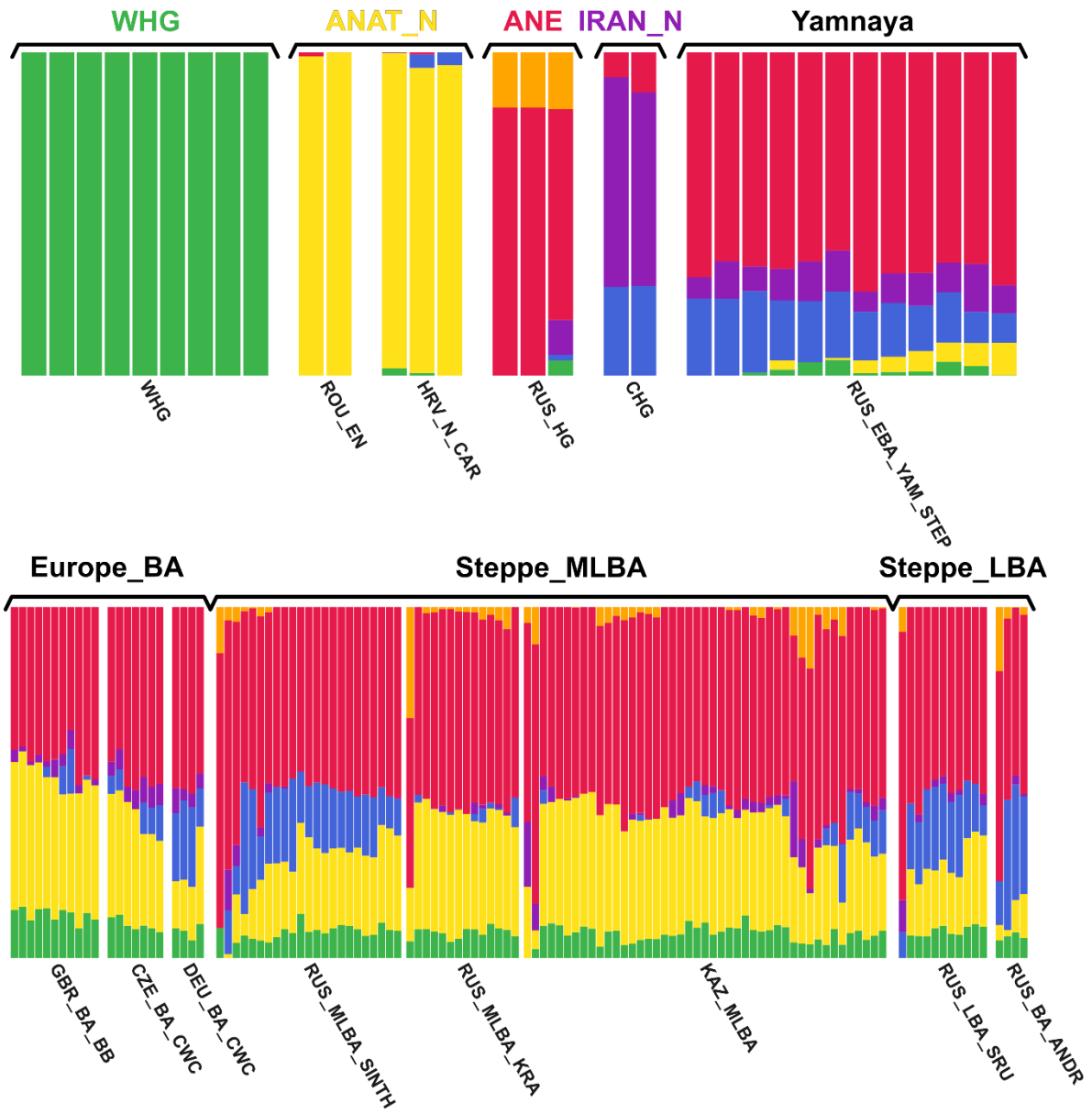


Figure 1: Unsupervised ADMIXTURE profile of key ancestral metapopulations from Eurasia and from the Bronze Age Steppe based on our own analysis (**Table E2**, for further discussion see **III.2.2** and **IV.3.2**). For details on the Short Label abbreviations also see **Table S5**.

Originally it was thought that the Early Bronze Age Yamnaya expansion to the West was paralleled by a similar event towards the East, as attested by the identification of the Yamnaya descendant Afanasievo culture in the Minusinsk Basin (Allentoft et al., 2015; Mathieson et al., 2015). However, this proved to be a sporadic migration and further data showed that the most likely candidate for a large-scale eastward migration event are the peoples of the Early-Middle Bronze Age Corded Ware culture (see Europe_BA on **Figure 1**; De Barros Damgaard et al., 2018; Narasimhan et al., 2019).

The eastward migrating groups came into contact with the previous inhabitants of the Steppe (possibly hunter-gatherers of the ANE population) and the great sedentary cultures of ancient Persia and of the Central Asian oases (Bactria-Margiana Archaeological Complex, BMAC). These genetic components were integrated on the fringes of the Steppe, but the dominant genetic composition remained remarkably similar throughout the whole of the Steppe region (see Steppe_MLBA on **Figure 1**) finalizing in the Andronovo archaeological horizon in the East and the Srubnaya culture on the Western Steppe (see Steppe_LBA on **Figure 1**). In the Late Bronze Age however, clear signs of a new genetic component of East Asian origin appears (ANA, orange component on **Figure 1** and **2**), as possibly climactic, economic and cultural changes drove a new epoch of migration waves from the East pointing to the West (Narasimhan et al., 2019).

I.2.2 The genetic makeup of the Iron Age nomads

From a backdrop of this widely distributed more-or-less homogeneous genetic background emerged the Scythians and the Sarmatians in the Early Iron Age (Gnecchi-Ruscone et al., 2021). In contrast to the Sarmatians, quite a lot of archaeogenetic discussion revolved around the origin of the Iron Age Scythians (134 samples published in 9 articles; the most important of these Gnecchi-Ruscone et al., 2021; Järve et al., 2019; Krzewińska et al., 2018; Unterländer et al., 2017). Actually, most of the publicly available samples identified as Sarmatians come from articles mainly concerned with the origin of the Scythians. This is understandable, as archaeological and historic data both points to a strong connection between these two groups, although their exact relation is still not resolved. As our main concerns did not involve the origin of the Scythians, it is enough to say that most articles agree in a scenario of independent multiregional emergence of the Scythians as a rather cultural phenomenon that connected East and West with limited gene-flow (Gnecchi-Ruscone et al., 2021; Järve et al., 2019). The Scythians of the Western Steppe (including the "Royal Scythians" of the Pontic region and Scythians from the Carpathian Basin) showed a distinct genetic composition built upon the Srubnaya-like genetic background, much more similar to the earlier Bronze Age groups of the region with strong ancestral patterns derived from Western Hunter Gatherers (WHG) and the Early Bronze Age Yamnaya people see (Western Steppe Scythians on **Figure 2**; Järve et al., 2019; Krzewińska et al., 2018). In contrast to this, the Scythians of the Eastern Steppe (individuals from the Aldy-Bel and the Pazyryk cultural sites as well as the Central Asian Sakas) showed the emergence of East Asian and Siberian-like genetic affinity (ANA, orange on **Figure 2**), which occurred only sporadically earlier in their region (see Eastern Steppe

Scythians on **Figure 2**; Gnecchi-Ruscone et al., 2021). Thus, two population centres seem to emerge in the literature, one in the West around the Pontic-Caspian and Southern Ural regions, and one in the East around the Altai Mountain range.

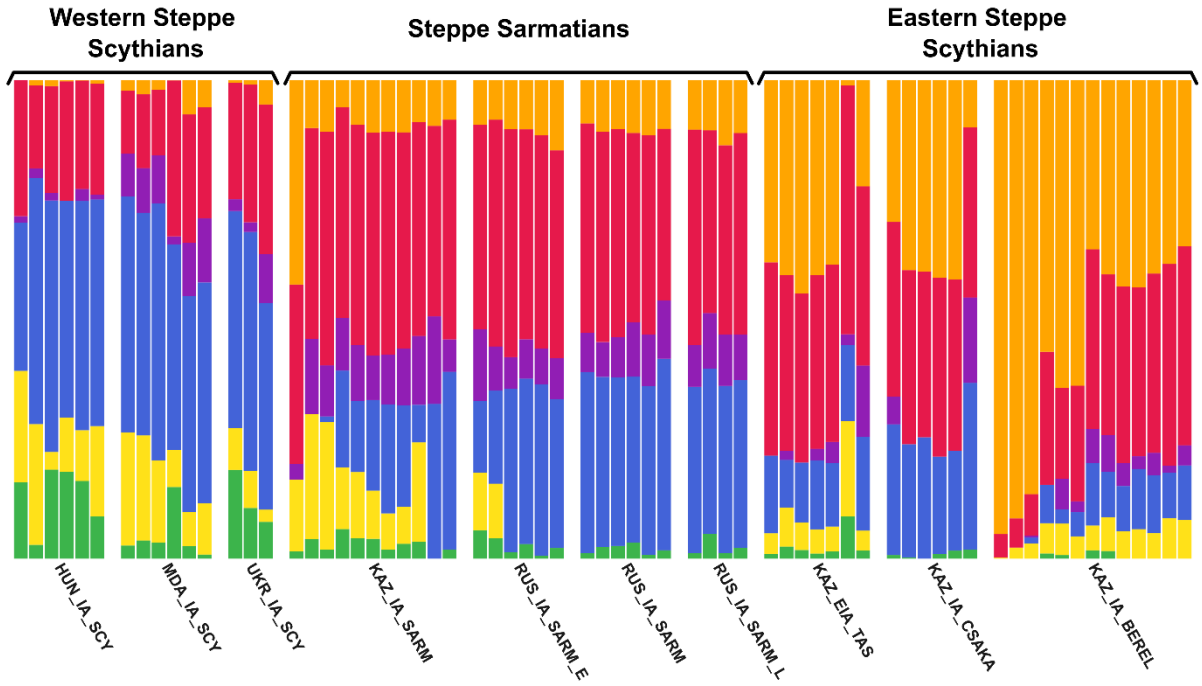


Figure 2: Unsupervised ADMIXTURE profile of Iron Age Steppe groups based on our own analysis (Table E2, for further discussion see **III.2.2** and **IV.3.2**). For details on the Short Label abbreviations also see **Table S5**.

I.2.3 Formation of the Steppe Sarmatians

The Sarmatians of the Steppe (henceforth referred to as **Steppe Sarmatians**) formed around the same time as the Scythians of the Western Steppes. Their area of origin may very well have been the Southern Ural region as attested by archaeological and genetic evidence (Istvánovits & Kulcsár, 2017). Before the 4th century BC, they only exerted a minor influence on the historical records, however around the 3rd century BC they become increasingly aggressive and begin to put strong pressure on the Scythian inhabitants of the Western Steppe region. By the turn of the millennium, they emerged as the dominant power on the Pontic-Caspian steppe to the point of becoming the new synonym for the steppe nomads in the eyes of their Roman neighbours. Ancient authors designated multiple tribes under their name which is also corroborated by numerous archaeological findings all around the North-Caspian and Central Steppe regions.

Despite this there are only a handful of articles in the literature referencing these influential people. The Allen Ancient DNA Resource (v54.1, Mallick et al., 2024) – which is quite up-to-date in this regard – only contains 45 definitive Sarmatian individuals from 7 different articles (Damgaard et al., 2018; Gnecchi-Ruscone et al., 2021; Järve et al., 2019; Krzewińska et al., 2018; Narasimhan et al., 2019; Unterländer et al., 2017; Veeramah et al., 2018). Among this only two articles discuss the Sarmatians in context (Gnecchi-Ruscone et al., 2021; Krzewińska et al., 2018), while the others only marginally. Because of the sporadic publishing of the available samples (3-5 genomes at a time), no articles provided definitive conclusion on their genetic composition and origin, as the low sample numbers only allowed for limited analyses.

General conclusions however can be collected from the literature. A unanimous agreement is apparent among the authors concerning the distinct genetic homogeneity of the Steppe Sarmatians which persist throughout a very wide spatio-temporal range (Gnecchi-Ruscone et al., 2021; Järve et al., 2019; Krzewińska et al., 2018). Furthermore, they exhibit the most pairwise genomic mismatch rate among the dominant Iron Age steppe groups (Krzewińska et al., 2018). According to the author, this points to a large effective population size, which together with the very apparent homogeneity of the Sarmatians may suggest a genetic continuity of the group in the southern Ural region despite a distinct cultural shift between the Early and Middle/Late Sarmatian Periods (Sauromatian-Sarmatian archaeological eras; Krzewińska et al., 2018). As they occupy an intermediate position between the Western and eastern Scythians on the PCA (and according to our ADMIXTURE analysis, see **Figure 2**), the author hypothesizes that they were a key intermediary in the demic diffusion of the Scythian culture, moreover the Southern Ural region may have been the original focal point in the formation of the Western Scythian base population (Krzewińska et al., 2018).

Järve and colleagues point out that a much higher affinity towards East Asian proxies can be observed in the Sarmatians compared to the previous populations of the region (e.g. individuals from the Sintashta or Srubnaya cultures; Järve et al., 2019). This is also observable on our own ADMIXTURE analysis (see Steppe Sarmatians on **Figure 2**). The presence of the elevated East Asian affinity well distinguishes the Steppe Sarmatians from their western Scythian neighbours, who seldom exhibit this genetic affinity. Furthermore, this affinity could point to stronger connection between the Southern Ural region and the Central Steppe nomads, however this relation is not explored substantially in the literature (only in Gnecchi-Ruscone et al. (2021) to a certain extent).

The Sarmatians - as other Iron Age nomad groups – show limited genetic affinity toward a South Eurasian source (purple on **Figure 2**). Gnecchi-Ruscone and colleagues point to an origin of this component in Turan as other proxies from Iran and the Caucasus consistently failed in their analysis. This again points to an existing connection between Central Asia and the homeland of the Steppe Sarmatians.

To summarize the available data:

1) The Steppe Sarmatians most probably formed around the Southern Ural region from the previous Steppe Middle-Late Bronze Age metapopulation (possibly Srubnaya). However, their genetic composition cannot be explained from this "Steppe composition" alone.

2) They exhibit remarkable genetic homogeneity, which persists through a wide geographic are (from the foothills of the Carpathians to the Kazakh steppe). This is coupled with a possibly large effective populations size, which suggests genetic continuity throughout the independent existence of their culture.

3) They harbour elevated proportions of East Asian and South Eurasian related ancestries linking them to the Eastern Scythian groups rather than their western neighbours.

II OBJECTIVES

Archaeological research on the Sarmatian Period in Hungary dates back to the early 20th century. The vast number of findings, including burials and settlement traces, highlight not only the cultural significance of the Sarmatians but also their potential impact on the population history of the Carpathian Basin. The extensive body of archaeological literature has enabled us to conduct high-resolution sampling across distinct periods associated with Sarmatian occupation. We collected a substantial number of samples from well-documented Sarmatian-period burials across a wide range of cemeteries in the Great Hungarian Plain, spanning the entire Sarmatian Period in the Carpathian Basin.

We employed traditional genome-wide variance analysis to address broad questions. However, advancements in genomic methodologies allowed us to explore genetic relatedness at an unprecedented level. By utilizing a refined IBD (identity-by-descent) detection method (Schütz et al., 2025), we identified distant genealogical relationships with high confidence, even revealing unexpected yet likely authentic close kinship ties between individuals separated by vast continental distances.

In summary, our main goals were as follows:

1. Characterize the genetic composition of the Sarmatian-period population of the Carpathian Basin.
2. Investigate the relationship between the Iron Age steppe nomads and the people found in the Carpathian Basin identified as Sarmatians.
3. If a connection exists, try to reconstruct the migration path the Sarmatians took to reach the Carpathian Basin.
4. Assess the genetic impact of the Sarmatians during their occupation and in subsequent periods of the Carpathian Basin's population history.

III MATERIALS AND METHODS

III.1 Ancient DNA pipeline

III.1.1 Sample collection

To uncover the genetic origin of the Sarmatian-period population living in the Carpathian Basin we focused our efforts to identify and collect human remains from a wider spatio-temporal range.

Concerning the Carpathian Basin Sarmatian Period proper, we aimed to thoroughly analyse the earliest cemeteries (usually described in the archaeological literature as the "golden horizon") and collect a representative sample set from the later periods with emphasis on the widest possible geographical range. Thus, almost all of the cemeteries included in this study yielded partial sampling. This approach allowed us to inference region wide population genetic changes and to ascertain typical features concerning the general population.

As the historical literature strongly suggests a southern route (through the lower Danube valley) for the arrival of the Sarmatian tribes from the Western Steppe (Istvánovits & Kulcsár, 2017), it seemed an obvious aim to obtain samples from the lower Danube region outside of the Carpathian range. These samples would allow the possible verification of the proposed southern migration of the Steppe Sarmatians and the study of genetic links between the Sarmatians residing in the Carpathian Basin and on the Western Steppes. Fortunately, a fruitful collaboration with anthropologists and archaeologists from Romania allowed us to acquire multiple samples from Wallachia and Moldavia (historical regions of Romania).

Finally, we thought it necessary to examine the afterlife of the Carpathian Basin Sarmatians, as historians and archaeologists are unsure how the Migration Period affected the populations of the region (Istvánovits & Kulcsár, 2017). To this end we also collected a representative sample from the immediately following Hun Period of the Carpathian Basin. This set helps to ascertain the possible survival of the Sarmatian-period population and also to contextualize the composition of the Migration Period population.

We collected and sampled 244 human remains in total. Following the archaeogenetic preparation of the samples we were able to obtain whole genome sequences in 156 cases. The distribution of the samples into the three greater categories can be observed in **Table 1**. Most of the unsuccessful preparations resulted from low endogenous content.

Table 1: Distribution of samples successfully and unsuccessfully processed by categories.

Category	Total	Successful
Outside of the Carpathian Basin	41	17
Sarmatian Period	158	118
Hun Period	45	21
All	244	156

We complemented this sample set with 17 Sarmatian-period individuals originally published in Gnecchi-Ruscone et al. (2022) and 9 Hun-period individuals from Maróti et al. (2022). As the samples from the former paper were generated by genome wide capture enrichment, we could not include them in all of the downstream analyses (namely imputation and IBD analysis). Nevertheless, we reanalysed them to the extent their coverage enabled us. Finally, we obtained 182 whole genomes resulting in the most comprehensive genome database representing the population of the Carpathian Basin between the 1-5th centuries AD.

III.1.2 DNA preparation

All steps of sampling, DNA extraction and library preparation were carried out as described in the Supplementary materials of (G. I. B. Varga et al., 2023), in the joint, dedicated ancient DNA laboratory of the Department of Archaeogenetics, Institute of Hungarian Research and the Department of Genetics, University of Szeged.

The bone samples were prepared with a minimally invasive extraction method described in (Harney, Cheronet, et al., 2021). We primarily collected teeth samples, where it was feasible and their connection to other bones of the studied individual could be securely determined. In every other case we sampled the petrous bone. Bone pieces were prepared with care using a Dremel® 3000 multifunctional hand drill and powdered with a VWR™ Star-Beater ball grinder.

DNA extraction was carried out by cleaning and soaking the whole teeth or 200 mg bone powder in digestion buffer (0.45 M EDTA, 250 µg/ml Proteinase-K, 0.1% Triton X-100) for 72 hours on 48 °C, than binding the DNA on Qiagen™ MinElute DNA purification columns with freshly prepared binding buffer (5 M Guanidine hydrochloride, 90 mM Sodium acetate, 40% Isopropanol and 0.05% Tween-20). Elution was carried out with a standard TE buffer (1 mM EDTA, 10 mM TRIS-HCl). The advantage of the teeth extraction method lies in the fact,

that the tooth samples could be retrieved after the DNA extraction process and safely deposited back into their original position in the sampled skulls.

We prepared double stranded DNA libraries according to the protocol described in Meyer & Kircher (2010) with minor modifications. We applied partial UDG treatment to counteract the effects of extensive postmortem damage (PMD). The reaction mix was prepared as described in Rohland et al. (2015) containing 1X Tango Buffer (Thermo Scientific™), 100 µM dNTPs, 1 mM ATP and 0.03 U/µl USER enzyme with 30 µl sample DNA. We incubated the samples for 30 minutes than stopped the reaction with Uracil Glycosylase Inhibitor (UGI). The samples were prepared for adapter ligation with blunt-end repair using a mixture of T4 polynucleotide kinase (0.5 U/µl) and T4 DNA polymerase (0.1 U/µl) and incubated for 20 minutes on 25 and 15 °C. Following this, the samples were purified on Qaigen™ MinElute columns and eluted in 20 µl Elution Buffer (EB). Adapter ligation was carried out according to Meyer & Kircher (2010). We used universal P5 and P7 adapter molecules in a mixture of 1X T4 DNA ligase buffer (Thermo Scientific™), 5% PEG-4000, 1.25 µM adapter mix and 0.125 U/µl T4 DNA ligase with 20 µl purified sample DNA. We incubated the samples for 30 minutes on 22 °C followed by another round of DNA purification on MinElute columns. Finally, we carried out an adapter fill-in reaction with 1X ThermoPol® reaction buffer (NEB®), 250 µM dNTPs and 0.3 U/µl Bst polymerase large fragment with 20 µl sample DNA to fill out the partially single stranded adapters and correct any nucleotide errors remaining on one of the strands. We omitted preamplification and directly double indexed our libraries in a single PCR step with Accuprime™ Pfx Supermix (Invitrogen™), containing 10 mg/ml BSA and 200 nM indexing P5 and P7 primers, in the following cycles: 95 °C 5 minutes, 12 times 95 °C 15 sec, 60 °C 30 sec and 68 °C 3 sec, followed by 5-minute extension at 68 °C. The indexed libraries were purified on MinElute columns and eluted in 20 µl EB.

The libraries were than shallow sequenced on Illumina iSeq 100 platform to monitor their human DNA content. Selected libraries were deep sequenced on Illumina NovaSeq 6000 platform using paired-end sequencing method (2x150bp) following the manufacturer's recommendations.

III.1.3 Data processing and quality control

The adapters of paired-end reads were trimmed using the Cutadapt software (Martin, 2011), and sequences shorter than 25 nucleotides were removed. Read quality was assessed with FastQC (Adrews, 2023). The raw reads were aligned to the Genome Reference Consortium

Human Build 37 (hs37d5) using the Burrows-Wheeler-Aligner (v 0.7.17) software, with the MEM command in paired mode, with default parameters and disabled reseeding (Li & Durbin, 2009). Only properly paired primary alignments with ≥ 90% identity to reference were considered in all downstream analyses to remove high mapping quality exogenous DNA containing non-aligned overhangs. Samtools v1.1 was used for merging the sequences from different lanes and also for sorting, and indexing binary alignment map (BAM) files (Li et al., 2009). PCR duplicates were marked using Picard Tools MarkDuplicates v 2.21.3 ('Picard Toolkit', 2019). To randomly exclude overlapping portions of paired-end reads and to mitigate potential random pseudo haploidization bias, we applied the mergeReads task with the options "updateQuality mergingMethod=keepRandomRead" from the ATLAS package (Link et al., 2017). Single nucleotide polymorphisms (SNPs) were called using the ANGSD software package (version: 0.931-10-g09a0fc5; Korneliussen et al., 2014) with the "-doHaploCall 1 -doCounts 1" options and restricting the genotyping with the "-sites" option to the genomic positions of the 1240K panel (Fu et al., 2015; Haak et al., 2015; Mathieson et al., 2015).

Ancient DNA damage patterns were assessed using MapDamage 2.0 (Jónsson et al., 2013), for the PMD (Post Mortem Damage) data see Schütz et al. (2025). Mitochondrial genome contamination was estimated using the Schmutzi algorithm (Renaud et al., 2015). Contamination for the male samples was also assessed by the ANGSD X chromosome contamination estimation method (Rasmussen et al., 2011), with the "-r X:5000000-154900000 -doCounts 1 -iCounts 1 -minMapQ 30 -minQ 20 -setMinDepth 2" options.

The raw nucleotide sequence data of the samples were published in Schütz et al. (2025) and deposited to the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession number: PRJEB80732.

III.1.4 Radiocarbon analysis

A high ratio of the burials from the Sarmatian and Migration Periods have been looted, and many of the analysed individuals came from partially excavated or solitary graves. To ensure accurate chronological placement, we conducted thorough radiocarbon dating of our samples. This was particularly crucial for individuals collected from outside the Carpathian Basin, where the Sarmatians occupied the region for a considerably longer period. In many cases, insufficient archaeological data made it impossible to determine their age without radiocarbon analysis.

We conducted radiocarbon analyses on 68 individuals. The measurements were done by accelerator mass spectrometry (AMS) in the AMS laboratory of the Institute for Nuclear

Research, Hungarian Academy of Sciences, Debrecen, Hungary (according to the methodology described in Molnár et al., 2013). The conventional radiocarbon dates were calibrated with the OxCal 4.4.4 software (https://c14.arch.ox.ac.uk/oxcal/OxCal.html, date of calibration: 20.02.2024) with IntCal 20 settings (Reimer et al., 2020). The analyses were mainly performed on skeletal bone fragments, however in viable cases part of the remaining petrous bone powder was used to limit the destruction of the human remains.

In most of the cases, the archaeological and radiocarbon dating confirmed each other. However, five samples showed major discrepancies between the two dating methods, with the radiocarbon results suggesting significantly older (earlier) dates (by more than one century). After ruling out issues with the archaeological dating and the possibility of sample misidentification, we suspected methodological issues with the radiocarbon analysis. Specifically, the much earlier dates may have occurred due to reservoir effects, such as the freshwater reservoir effect linked to a fish-based diet (Jull et al., 2013).

The underlying cause of this phenomenon lies in the significantly lower natural $^{14}C/^{12}C$ ratio in marine environments, which results from the dissolution of ^{12}C from waterlogged mineral deposits. Marine ecosystems, characterized by complex trophic chains, tend to incorporate and concentrate this "old" carbon to a greater extent than terrestrial ecosystems. Consequently, the radiocarbon dating of marine organisms - and of carnivores consuming substantial amounts of marine-based food - routinely yields dates that appear significantly older than their actual age. A similar, though typically less pronounced, effect occurs in freshwater environments (Jull et al., 2013).

For humans, high levels of fish consumption can also lead to erroneous radiocarbon dates that are artificially earlier by several hundred years (Philippsen, 2013). The reservoir effect can be mitigated by analysing land-based herbivore remains found in the same grave context and comparing their radiocarbon dates with those of the suspected individuals. Furthermore, $\delta^{13}C/\delta^{15}N$ stable isotope ratios can be used to identify elevated fish consumption, indicating a potential reservoir effect. However, as isotopic signatures are influenced by multiple environmental and dietary factors, herbivore reference measurements may still be required for accurate interpretation (Ramsey et al., 2014).

Unfortunately, no background data - such as from domesticated animal bones - was available to help filter or quantify the impact of the reservoir effect. To address this, simulations were performed to estimate the potential impact of varying levels of freshwater-related diets (see **Extended Figure 1**).

As with the calibration, we used the OxCal 4.4.4 software to perform the simulations (https://c14.arch.ox.ac.uk/oxcal/OxCal.html). First, an expected calibrated date distribution was generated using the "R_Simulate()" function, based on archaeological dating estimates. Next, a prior distribution was produced with the "R_Date()" function, derived from the actual radiocarbon measurements of the samples. Finally, posterior distributions were created using the "Mix_Curves()" function, assuming various levels of mixed marine and atmospheric diets (20%, 30%, and 40%). The simulation approximates real dietary effects when the posterior distribution maximally overlaps with the expected date range (presuming the archaeological dating was accurately assessed).

III.1.5 Haplogroup assignment

Mitochondrial haplogroups were determined using the HaploGrep 2 (version 2.1.25) software (Weissensteiner et al., 2016), using the consensus endogen fasta files resulting from the Schmutzi Bayesian algorithm. The Y chromosome haplogorup assessment was performed with the Yleaf software tool (Ralf et al., 2018), updated with the ISOGG2020 Y tree dataset. In one case (MDH-405) the Y Hg could not be determined due to low coverage, this was marked as "inconclusive".

III.1.6 Map drawing

Maps were created in R 4.1.0 (R Core Team, 2018) with the help of "ggplot2", "sf", "ggspatial", "rnaturalearth" and "elevatr" packages (Dunnington, 2023; Hollister et al., 2023; Massicotte & South, 2024; Pebesma & Bivand, 2023; Wickham, 2016). First, a data table was called containing the spatial coordinates for the selected geographical area using "rnaturalearth". Waterways were added as a separate layer using the spatial data published in Yan et al. (2022). Finally, we obtained elevation data with "elevatr" and superimposed it on the spatial coordinates as a "geom_tile" using "ggplot2".

III.2 Analysis methods

III.2.1 PCA

To uncover the underlying structure of our studied individuals in a hypothesis-independent manner, we conducted Principal Component Analysis (PCA). PCA is an orthogonal, linear transformation that projects the original data points onto new synthetic variables (axes) that best capture their variance. Each new axis is a weighted linear combination of the original variables, with the first principal component (PC1) accounting for the highest variance, the second principal component (PC2) capturing the next highest variance, and so on.

In this analysis, genetic data is structured as a matrix, where individuals form the rows and single nucleotide polymorphism (SNP) loci form the columns. Since we consider only biallelic loci, each data point is encoded as either 0 (reference allele) or 1 (alternate allele) in a pseudo-haploid framework. The PCA process can be visualized as identifying a new axis – the first Principal Component (PC1) - that will maximize the distance between the points projected onto it. The second axis (PC2) will have the same property but with a constraint that it must be perpendicular (orthogonal) to the first axis. And the *k*th axis (PC*k*) will have the same property but must be orthogonal to the *k-1*st axis. In practice, PCs are calculated by solving the characteristic polynomial of the covariance matrix of the original data.

PCA offers several advantages over analysing raw genetic data directly. First, the evolutionarily neutral SNP set - usually under consideration - does not carry explicit meaning, rather it's allele distribution and variance exhibited among the studied individuals harbours its real value. Second, meaningful genetic differences emerge across the entire genome rather than at single SNP sites, requiring a method that captures genome-wide variation. PCA addresses both concerns by constructing new axes that summarize underlying genetic variance while aggregating information from all loci. This approach allows us to condense genomic data containing hundreds of thousands of SNPs into just two dimensions, maximizing the visual separation of individuals in a readily understandable way. Conceptually, PCA can be imagined as rotating an *n*-dimensional dataset (where *n* is the number of loci) to find the optimal 2D projection that preserves the greatest genetic differentiation.

According to the recommendations of Elhaik (2022), we constructed our genetic map by projecting our ancient individuals onto PC axes calculated from a large dataset of modern Eurasian individuals. While this method may distort genetic variations unique to historical populations, its key advantage is that it provides a robust and reproducible framework for

comparison. As background, we used the same modern Eurasian genome dataset as our previous publication (Maróti et al., 2022) confined to the HO SNP set. This consisted of a generalized set of 1397 modern individuals from 179 modern Eurasian populations (for the PCA distribution of the modern individuals see **Extended Figure 2**).

PCA Eigen vectors were calculated from these pseudo-haploidized modern genomes with smartpca (EIGENSOFT version 7.2.1, Patterson et al., 2006). All ancient genomes were projected on the modern background with the ‘‘lsqproject: YES and inbreed: YES’’ options. Since the ancient samples were projected, we used a more relaxed genotying threshold (>50k genotyped markers) to exclude samples only where the results could be questionable due to the low coverage.

III.2.2 ADMIXTURE

We used ADMIXTURE analysis to model our genomes as compositions of hypothetical ancestral populations (Alexander et al., 2009).

ADMIXTURE uses expectation maximalization to estimate the parameter values of a predictive model, which best describe the observed SNP states of our genetic data. Its underlying model describes hypothetical ancestral populations with allele frequencies *f* that produce descendant populations with a *q* proportional contribution. The likelihood of a set of descendant individuals to inherit their allele frequencies from *K* ancient populations with *f* allele frequencies contributing *q* proportions is equal to the probability of the descendants to exhibit their observed allele frequencies given a specific set of *f* and *q* parameters. During the calculation process we are trying to iteratively find that parameter combination which maximizes this probability (maximum likelihood estimation). The caveat is that the appropriate parameter combination must accurately predict all of the observed allele frequencies across all individuals and loci at the same time. To obtain this in a reasonable runtime, Alexander and his peers had to integrate multiple optimization processes like block relaxation and a novel quasi-Newton method for accelerated convergence (Alexander et al., 2009).

We begin with a dataset of individuals with an observable intersecting SNP set. When we perform the analysis, we have to give the number of hypothesised ancestral populations (*K*) than the maximum likelihood estimation calculates the best fitting *q* contribution coefficients of the *K* ancestral populations for each individual and the *f* allele frequencies of these ancestral populations. We are mainly focusing on the obtained *q* values, as these can be useful in multiple ways. First, we obtain an ancestral component composition for each individual in a hypothesis

independent manner. It can be assumed that individuals with a similar genetic background have to display very similar component composition. Furthermore, individuals truly descending from the same ancestral metapopulations should also share this disposition in their ADMIXTURE profile. Second, although the analysis actually calculates the inferred allele frequencies of the hypothetical ancestral populations it might as well be that the true ancestral populations are present in the analysed dataset. In this case it can be reasonably assumed that the algorithm will correctly identify these individuals and confer to them an appropriately large component fraction. This way real genetic meaning can be associated with the obtained ancestral components. Unfortunately, there are some limitations to this method, as model convergence requires for each individual to obtain some results. Thus, genetic outlier individuals, scarcely represented in the data may obtain unreliable component values. A further consideration has to be given to close relatives, as they realistically share long parts of their genome, thus the analysis may erroneously identify these SNP sets as indicators of stratified ancestral history.

We performed unsupervised ADMIXTURE analysis on 2578 ancient individuals using the 1240K SNP set (Mallick et al., 2024). This included **149** newly sequenced individuals and **2429** ancient individuals from the Allen Ancient DNA Resource (AADR; Mallick et al., 2024). As we did not include modern individuals from the HO dataset, we obtained a significantly higher number (391,178) of overlapping SNP sites. We pruned variants in linkage disequilibrium using PLINK (Chang et al., 2015) with the options ‘‘–indep-pairwise 200 10 0.25’. We excluded individuals with less than 200K SNPs covered and samples with >4% contamination, we also taken out close relatives to avoid the appearance of undesired ancestral components. The results for this analysis can be found in **Table E1**. Cross-validation error calculation showed that modelling K-6 ancestral components yielded the most consistent results (**Table E1b**).

III.2.3 F4-statistics

F4-statistics allow us to infer allele frequency covariance among four populations, providing a hypothesis test for specific population tree structures. The statistic is calculated as the product of the allele frequency differences between two population pairs. The results are summed across all loci and then averaged by the number of analysed sites. If the final value is not significantly different from zero, it indicates that at least one of the population pairs shares a high degree of genetic homogeneity. In such cases, the proposed tree structure is necessarily valid, as F4-statistics always assess a rootless tree.

A specific way of representing a tree structure as well as a method for testing single individuals was introduced by Green et al. (2010) under the name "D-statistics". This approach incorporates an outgroup that is symmetrically related to the tested individuals, allowing the detection of directional gene flow from a given individual into two reference populations. Patterson et al. (2012) later demonstrated that F4-statistics and D-statistics describe the same underlying test statistic but from different theoretical perspectives. Consequently, these methods have been integrated into a unified analytical framework capable of assessing treeness (the fit of populations to a bifurcating tree model) at both the individual and population levels.

In our analysis, we employed two types of rooted trees based on the D-statistics approach of Green et al. (2010). First, we tested for simple directional gene flow or genetic affinity using the formula **F4(Outgroup, Test; Ref1, Ref2)**. Here, the Test represents the studied sample, while Ref1 and Ref2 are reference populations. A negative F4 value indicates a stronger affinity between the Test and Ref1, while a positive value suggests a closer relationship to Ref2.

Second, we leveraged the exclusive nature of F4-statistics to our advantage. The test **F4(Outgroup, Ref1; Ref2, Test)** only produces a meaningful value at loci where both the Outgroup-Ref1 and Ref2-Test pairs exhibit different alleles. This approach is particularly useful because F4-statistics measure net genetic similarities, making them less effective in detecting complex, overlapping affinities. If the Test has a strong genetic relationship with one reference population, it may obscure weaker affinities with the other. By subtracting the major affinity, we can reveal minor affinities that would otherwise remain undetectable using the more conventional **F4(Outgroup, Test; Ref1, Ref2)** approach.

We tested multiple F4-statistic frameworks (see **III.2.3**). The statistics were calculated using the 1240K SNP set with the qpF4ratio algorithm from ADMIXTOOLS (Patterson et al., 2012). The results of the F4 analyses were visualized in a 2D framework.

III.2.4 qpAdm

We used qpAdm (Harney, Patterson, et al., 2021) from the ADMIXTOOLS software package (Patterson et al., 2012) for modelling our genomes as admixtures of two or three source populations and estimating ancestry proportions. The qpAdm analysis was done with the HO dataset, as in many cases suitable RIGHT or LEFT populations were only available in this dataset.

Our main goal of the study was to investigate the relationship between the Sarmatian individuals found in the Carpathian Basin and Sarmatians of the Central Steppe. Thus, in our qpAdm

analysis framework we wanted to evaluate whether the available steppe Sarmatian individuals are a necessary source population for modelling the studied individuals. We considered three types of possible modelling sources (LEFT populations); a) a population set representing the supposed local inhabitants of the region (indicated by grey in **Table E2**), b) a population set representing the proposed Sarmatian ancestry (indicated by green in **Table E2**), and c) a population set representing other possible Central and East Asian sources (indicated by light orange in **Table E2**) as our ADMIXTURE analyses indicated at least a marginal appearance of these.

The set of 15 "local" sources were selected from an extensive preliminary qpAdm run, among 162 possible candidates from the AADR (Mallick et al., 2024). We performed all possible 2-way combinations without model competition, with the primary aim of narrowing down the LEFT population set to optimize runtime for the subsequent model competition. Observations from this preliminary analysis were used to assess the adequacy of candidate sources. We also evaluated the PCA clustering of all proposed source individuals to avoid redundant sampling, as well as ADMIXTURE component profiles to ensure the representation of components found in our Test individuals. Final source populations were selected based on sufficient genome coverage and fulfilment of at least one of the following criteria: a) the source was present in all acquired models, b) the sources represented a unique cluster in the PCA space, c) the source exhibited appropriate ADMXITURE components (did not contain an additional ancestral component that was not present in the Test individuals or a component that was not present in the Steppe Sarmatians and its ratio was smaller than the Test individuals').

We assembled this optimal source population subset for the explicit aim to model the highest number of our test subjects in a single qpAdm run and not to acquire their true ancestral composition. To represent the arriving Sarmatian population we assembled two genetically homogeneous source population from the available Steppe Sarmatians published from Russia and Kazakhstan (Mallick et al., 2024). The remaining sources represent other possible Central or Eastern Asian immigrants. We also used populations from our previously published article (Maróti et al., 2022), which had been shown to have extensive connections to the Carpathian Basin.

The reference population set (RIGHT populations) contained Ethiopia_4500BP (fixed), Iran_GanjDareh_N, Turkey_N, Latvia_HG, Baikal_EN (Russia_Shamanka_Eneolithic.SG and Russia_Lokomotiv_Eneolithic.SG), WSHG (Russia_Tyumen_HG and Russia_Sosnoviy_HG),

Russia_Steppe_Maikop, Karitiana and Poland_Koszyce_GlobularAmphora.SG. For detailed list of LEFT and RIGHT populations see **Table E2**.

During the runs we set the details:YES parameter to evaluate Z-scores for the goodness of the fit of the model (estimated with a Block Jackknife). As qpWave is integrated in qpAdm, the nested p values in the log files indicate the optimal rank of the model. This means that if p value for the nested model is above 0.05, the Rank-1 model should be considered (Harney, Patterson, et al., 2021).

As our extensive source population set (LEFT populations) portended a great number of alternate models we applied the model competition framework explicitly discussed in Narasimhan et al. (2019) and Maróti et al. (2022). In this setup we test each resulting qpAdm model with a feasible p-value, by iteratively rerunning it with moving each of the LEFT populations in the RIGHT population set. This results in a bilateral improvement. On one hand the true source population – when included in the RIGHT population set – should consistently exclude any suboptimal models, as the test will (by design) have their highest shared drift with their true sources. On the other hand we will have a distribution of p-values for each resulting model which enables us to better quantify their goodness of fit, as demonstrated in Harney, Patterson, et al. (2021). As we run each model multiple times, we can obtain further useful information concerning the feasibility of the individual models. Thus, based on the output of the qpAdm algorithm we included further quality measurements into our analysis framework.

In light of recent research on the robustness of the qpAdm framework, we must acknowledge its potential for false positive results, particularly when analysing highly complex parameter fields. For a detailed discussion of this topic see Flegontova et al. (2025). Given these limitations, interpretation of high complexity qpAdm analyses should be approached with caution and restricted to broad, general conclusions. Nevertheless, we believe our results remain valid, as their credibility is supported by the joint interpretation of multiple independent analyses results as well as the intentionally modest aims of our qpAdm approach.

The meaning of the columns in **Table E2a-c** is as follows. **Test**: the name of the individual/population modelled. **SourceX**: the name of the designated sources in descending order of contribution. **SourceX ratio**: the obtained mixture coefficient for the source population/individual in question. **Valid models**: the number of cases a given model passed the quality criteria. **Excluded models**: the number of cases a model has been excluded by one of the reference populations. **BadFit models**: quality criterion, number of cases the obtained mixture coefficient differs from the jackknife calculated mixture coefficient by at least 10%.

Negative models: quality criterion, number of cases where one of the source coefficients is lower than 0. **Non-significant nested p-value models**: the number of models where the obtained nested p-value is higher than 0.05. This indicates that the k-1 model is more plausible. **Average nested p-value**: the average of the obtained p-values for the k-1 model. **Minimum p-value**: the lowest p-value obtained from the model competitions. **Maximum p-value**: the highest p-value obtained from the model competitions. **Average p-value**: the average of the obtained p-values of the valid model repetitions (passing the quality criteria). **Average p-value summary**: either the average p-value or the average nested p-value if the number of non-significant nested p-value models is higher than the half of the valid models. This was used to order the qpAdm results. **Minimum p-value reference**: the name of the LEFT population/individual which was moved to the RIGHT population set when the lowest p-value was obtained for the model in question. **Maximum p-value reference**: the name of the LEFT population/individual which was moved to the RIGHT population set when the highest p-value was obtained for the model in question. **Excluding reference**: the name of the LEFT population/individual which was moved to the RIGHT population set when the model was excluded. **Used on qpAdm plot**: an "x" signifies the specific model that was used to compile **Figure 10**.

We ran comprehensive 2-way modelling runs for all studied TEST individuals with the above-described RIGHT population set and freely combined LEFT population set (Table E2a). Subsequent 3-way modelling was only conducted on a selected subset of the TEST individuals with unsatisfactory 2-way models (Table E2b). We selected these individuals based on preliminary qpAdm analyses where a sufficient increase in their p-value was reasonably expected. After 3-way modeling, 12 individuals still remained with no feasible models (p-value <0.05 or all models excluded by the model-competition). As the ADMXITURE and PCA profile of these individuals showed a very similar composition as outlier individuals in our previous article (Maróti et al., 2022), we included some further sources representing possible Northern (indicated by light blue in **Table E2c**) and Southern European (indicated by gold in **Table E2c**) populations of the time as some of the outliers seemed clearly deriving a portion of their ancestry from these regions. We swapped our Steppe Sarmatians sources for some of our already modelled individuals and individual Russian and Kazakh Sarmatians, as they may contain some minor components that were not sufficiently represented in the grouped Sarmatian reference populations. In the end we successfully modelled all of our studied individuals except

for a single individual (HVF-10) with high average p-values. The single unmodeled individual seems to be an outlier which has no sufficient source in the database yet.

III.2.5 Imputation

We imputed our studied genomes together with other shotgun sequenced ancient genomes from a similar spatio-temporal distribution with the GLIMPSE2 framework (version 2.0.0; Rubinacci et al., 2021) according to the recommendations of Sousa da Mota et al. (2023). Approximately 78 million biallelic common markers from the 1KG dataset were imputed with GLIMPSE2, utilizing the 1KG phase III data as a reference. The reference dataset was normalized, and multi-allelic sites were split using bcftools (version 1.16-63-gc021478 with htslib 1.16-24-ge88e343), applying the "norm –m –any" subcommand. Biallelic SNPs were filtered using the "view –m 2 –M 2 –v snps" subcommand. The autosomal chromosomes of the human reference genome were divided into 580 genomic chunks using the GLIMPSE2_chunk tool with the "-sequential" option. Following the GLIMPSE2 guidelines, we generated the binary reference data using the GLIMPSE2_split_reference tool, based on the 580 genomic regions and 1KG biallelic SNP variants. For the imputation process, we included only samples with shotgun WGS data exceeding 0.25x mean genome coverage and mitochondrial or autosomal contamination levels below 0.04, as recommended in the GLIMPSE2 manuscript. Due to these criteria, we excluded 17 Sarmatian samples published in Gnecchi-Ruscone et al. (2022), as they were obtained through capture enrichment sequencing, along with 7 of our newly sequenced genomes that exceeded the contamination threshold.

III.2.6 Kinship estimation and IBD analysis

Kinship analysis was performed with correctKin (Nyerki et al., 2023). As reference population we applied the same database as in G. I. B. Varga et al. (2023). The results for the kinship identification are present in **Table S4**.

The shared IBD segments were identified using the ancIBD framework (version 0.5; Ringbauer et al., 2024) according to the recommended workflow outlined in the official ancIBD documentation (https://ancibd.readthedocs.io/en/latest/run_ancIBD.html). Phased and imputed variants were post-filtered to include only the positions of the 1240K AADR marker set and lifted to the hdf5 data format using the 'vcf_to_1240K_hdf' method with the default parameters. IBD fragments were identified using the 'hapBLOCK_chroms' method applying the standard five-state HMM with the haploid_gl2 emission model, which is appropriate for the GLIMPSE2 posterior likelihoods, to ascertain the raw IBD segments >=8cM.

During the subsequent filtration of raw IBD segments, we deviated from the marker density threshold (≥ 220 SNPs/cM) used in the original ancIBD framework (Ringbauer et al., 2024). We found that applying this global threshold led to a significant number of false positive and false negative IBD segment identification. Instead, we implemented a novel method that uses marker informativity scores to dynamically mask and exclude genomic regions lacking sufficient power to detect true IBD segments (for further details see Schütz et al., 2025). We developed this filtration method in Python for use with the raw IBD output generated by ancIBD. Our tool integrates seamlessly into the ancIBD framework and is available on GitHub [www.github.com/zmaroti/scoreFilterIBD].

Shared IBD network was generated in R 4.1.0 (R Core Team, 2018), with the application of packages ggplot2 3.4.2 (Wickham, 2016), igraph (Csárdi et al., 2024) and qgraph (Epskamp et al., 2012). The results of the IBD analysis can be found in **Table E3**.

III.2.6.1 Plotting

To generate **Figure 11A**, we partitioned our dataset into wider spatio-temporal groups based on the archaeological and geographical data. We first contracted our whole IBD network into single points (vertices) representing each group, then calculated the distribution of these points using the Fruchterman-Reingold weight directed algorithm implemented in qgraph (Epskamp et al., 2012), where weights were the number of connections (edges) between each group. We centralized and expanded this distribution to create sufficient space, then we calculated graph distributions for each groups separately with the same weight directed algorithm, but now the weights were given as the sum total length of shared IBDs between each individual. We added the coordinates obtained from the single vertex calculation to each corresponding groups' coordinates to arrange the separate group distributions according to the single vertex distribution. Finally we ran a new weight directed algorithm, with the recalculated coordinates as initial coordinates and with *niter = 1* and *max.delta = 0* parameters to not allow points deviating from their initial coordinates. We only plotted edges corresponding to intergroup connections to make the plot more straightforward. Intragroup connections are represented in this cause by the distance of the points from each other.

For **Figure 11B** we allowed the algorithm to run for 100 iterations and we used *max.delta = (0.1 + % between-group connection)* which allows for points to move along their edges but proportional to the number of intergroup connections. This approach emphasizes the net direction of intergroup attractions.

For **Figure 12** we defined a core set of populations that we were primarily interested in. This included individuals with the Regional Group labels (also found on **Figure 11**): Romanian Sarmatian, Carpathian Basin Sarmatian Period and Carpathian Basin Hun Period, which included the individuals published in this article, and Central Steppe Sarmatian, Carpathian Basin Avar Period, Carpathian Basin Conquest Period as main candidates for references (**Table E3**). We included a further 17 individuals from other Regional Groups that had at least five connections to any of the core populations as to not inflate the plot with non-informative data. This partition was made to find an equilibrium between a strict constrain that also preserves only the most informative individuals. After defining the desired individual set, we ran a simple Fruchterman-Reingold weight directed algorithm with no constrains for 1000 iterations, where weights were given as the sum total length of IBD fragments shared between individuals.

An important metric for graph-based calculations is the number of connections an individual has (degree or degree centrality, *d*). When comparing degree centrality among groups or individuals it's important to consider that a connection always exists between two endpoints, thus we carefully avoided counting individual connections multiple times. Another pitfall to consider is comparing groups of different sizes, where the sample size disproportionately affects the chance of finding IBD connections (linearly increasing sample sizes quadratically increase the possibility of uncovering connections) thus, to properly compare groups we must normalize the obtained degree centrality values by dividing them with the number of possibly available connection thus producing the ratio of fulfilled connections. For **Figures 13** and **14** the number of connections (degree centrality, *d*) was normalized by the product of the sizes of the two groups in case of intergroup connections ($d_i' = d_i / [n_i \times n_j]$), while intragroup connections were normalized by the equation: $d_i' = d_i / ([n_i \times \{n_i-1\}]/2)$, where *n* symbolizes the group size. **Figure 15** was prepared by analysing the number of connections between individuals of a specific cemetery and all other individuals from the Carpathian Basin Sarmatian and Hun Periods. We normalized the degree centrality of the cemetery groups by dividing it with the product of the size of the cemetery groups (n_i) and the number of all remaining individuals ($n-n_i$): $d_i' = d_i / (n_i \times [n-n_i])$.

III.2.7 Haplogroup data comparison

To shed light on the most feasible origin of the uniparental lineages of our samples, we assembled a comprehensive uniparental database of the Carpathian Basin. We have chosen samples from the AADR (Mallick et al., 2024) database based on their country of origin to cover this region. This included samples from Hungary, Slovakia, Romania, Serbia, Croatia,

Slovenia and Austria. The haplogroups were collected from the original publications and cross-referenced with the publicly available databases published in (Freeman et al., 2020; Maár et al., 2021) as well as our own classifications in the cases where the genomes were already downloaded for other analyses (**Table E4**).

IV RESULTS

IV.1 Sample data

IV.1.1 DNA sequencing results

In our pursuit of mapping the genetic landscape of the Sarmatian-period Carpathian Basin we successfully generated 156 shotgun sequenced whole genomes from a wide spatio-temporal range. The samples showed a mean coverage of 1.42-fold (0.24x-3.75x) with negligible contaminations (**Table S1**).

As mentioned in **III.1.1**, we supplemented our dataset with 26 previously published genomes from Gnecchi-Ruscone et al. (2022) and Maróti et al. (2022). Although their genetic data is detailed in their respective publications, we include a brief summary table in **Table S2**. Rather than reclassifying these samples, we retained their original labelling, which fortunately aligned with the classification scheme applied in this work.

In total, we collected 182 samples from 65 cemeteries with an average sample size of 2.8 per cemetery. This broad yet shallow sampling strategy enabled us to draw general conclusions about the Sarmatian-period population of the Carpathian Basin while avoiding overemphasis on local variations and specific case studies.

IV.1.2 Radiocarbon dating

The samples were thoroughly reviewed for precise archaeological classification, a detailed description of the analysed cemeteries can be found in Schütz et al. (2025). We also performed radiocarbon measurements in 68 cases to anchor and validate the archaeological dating approach (**Table S3**).

In most cases, radiocarbon and traditional archaeological dates confirmed each other, with only minor variance in their upper or lower ranges. We classified our samples mainly based on the existing archaeological periodization of the era. This included 4 Sarmatian phases: Early, Early-Middle, Middle-Late, and Late Sarmatian Period, as well as 2 phases from the Hun Period. However, five samples exhibited major inconsistencies between the radiocarbon and archaeological dates, with radiocarbon dates suggesting significantly earlier dates than the archaeological context would indicate. Analysing the genetic data from these individuals pointed toward potential methodological issues concerning the radiocarbon dating. Specifically

the possible presence of reservoir effect, often cased by elevated fish consumption, that can push back radiocarbon dating by hundreds of years (Jull et al., 2013).

As we lacked reference samples (e.g. domesticated herbivores from the same location) to directly confirm the presence of reservoir effect, we conducted simulations to assess the potential impact of a partially fish-based diet on our radiocarbon dating results (**Extended Figure 1**, for further details see **III.1.4**). These simulations indicated that fish consumption levels of 20–30% could account for the older radiocarbon dates, which is plausible given that all the samples came from the southern Great Hungarian Plain, near the Tisza and Maros rivers, key economic resources for local populations since ancient times. Consequently, in these cases, we prioritized the traditional archaeological dates over the radiocarbon dates.

Reservoir effect was particularly evident in one individual (NKL-7), whose archaeological age was ambiguous. Radiocarbon dating placed NKL-7 in the early 1st to 2nd century CE (84-95 (3.3%); 116-216 (92.2%) calCE), however, this early date is strongly contradicted by genetic evidence. NKL-7 shares multiple 4th-degree relations with unpublished Early Avar-period individuals from Tiszavasvári-Kashalom dűlő (TKD) and Üllő (ULL), based on kinship coefficient estimates (**Table S4**) and identity by descent (IBD) analysis, resulting in a discrepancy of approximately 300 years between the genetic and radiocarbon data. Therefore, the early radiocarbon date is likely due to a reservoir effect.

The same applies to two samples from the Óföldeák - Ürmös (OFU) cemetery (OFU-168, OFU-422). The radiocarbon dating for OFU-168 was 127-232 CE and 261-415 CE for OFU-422. Nevertheless, these individuals had multiple close familiar connections to Late Avars from the unpublished Tiszafüred – Majoroshalom (TMH) site (**Table S4**). Therefore, if reservoir effect is considered, these individuals may rather be dated to the early Avar Period, although their archaeological material does not explicitly point to this.

The samples obtained from Romania were thoroughly analysed with radiocarbon dating to verify their age. This proved to be necessary as two individuals (LMO-8 and RAK-7) turned out much older than the rest, dating to the Early Iron Age periods of Eastern Europe. These datings were also not contradicted by their archaeological description, which unfortunately was quite lacking otherwise. We include these individuals in the publication as Western Steppe Iron Age (ROU_IA).

IV.1.3 Sample classification

The classification of the Carpathian Basin Sarmatians is a complex issue with a substantial body of literature (detailed in **I.1.5**). While radiocarbon data was available from most of the cemeteries, we did not pursue a one-sided approach as no single method seemed to represent the best solution.

While radiocarbon dating might seem the most accurate approach for periodization, its precision over short time spans, such as the 100–150 years separating adjacent Sarmatian Periods, is uncertain and subject to significant variability. Reservoir effects, as detailed in the previous section (**IV.1.2**), can further complicate results, and detecting these would require extensive sampling of animal bones from the analysed sites that we did not perform. Furthermore, we were only able to perform radiocarbon measurements on a subset of our samples, as this analysis is very costly, with the expense of a single measurement comparable to that of whole-genome sequencing.

In contrast, archaeological data was available for most of the analysed samples. However, this approach also comes with its own challenges. The archaeological periodization is based on the gradual evolution of associated artifacts, where newer types can be derived from older ones, but not the other way around. Additionally, there are specific artifact types that can be accurately dated, such as coins and Roman pottery. While the lower boundaries of periods can often be established with reasonable confidence using these methods, the upper boundaries are more ambiguous. Furthermore, cultural transition boundaries are rarely fixed and can vary significantly, even between micro-regions.

Recognizing the limitations of both approaches we adopted an integrated periodization strategy. We did not rely solely on the radiocarbon results rather integrated verified archaeological data with radiocarbon dating. Samples were assigned to 8 separate groups when both independent lines of evidence confirmed their accurate classification (**Table 2**).

Table 2: Sample groups used in this study. The abbreviations termed “Short Label” are consistently applied throughout this work for all groups, following the format: three-letter country code _ archaeological age _ cultural label _ additional specific identifier.

Short Label	No. of samples	Archaeologocal context	Country of origin
ROU_IA	2	Western Steppe Iron Age	Romania
ROU_SARM	15	Western Steppe Sarmatian Period	Romania
HUN_SARM_EP	15	Carpathian Basin Early Sarmatian Period	Hungary
HUN_SARM_EMP	13	Carpathian Basin Early-Middle Sarmatian Period	Hungary
HUN_SARM_MLP	55	Carpathian Basin Middle-Late Sarmatian Period	Hungary
HUN_SARM_LP	22	Carpathian Basin Late Sarmatian Period	Hungary
HUN_SARM_UP	13	Carpathian Basin Sarmatian Unkown Period	Hungary
HUN_SARM_HUN	10	Carpathian Basin Late Sarmatian-Hun Period	Hungary
HUN_HUN	11	Carpathian Basin Hun Period	Hungary

The spatial and temporal distribution of our samples is illustrated in Figures 3A and 3B, respectively. As mentioned in the previous section, our groups were mainly constructed on an archaeological basis and represent more or less progressive periods (see **Figure 3B**). To prevent potential contradictions from inaccurate periodization, an "unknown" group was created for samples with radiocarbon results but lacking precise archaeological dating (HUN_SARM_UP, see **Table 2**). These samples were excluded from downstream group-based analyses. The samples from previous articles and the ones sequenced here were distinguished on some plots by the addition of a “PUB” suffix but otherwise handled together.

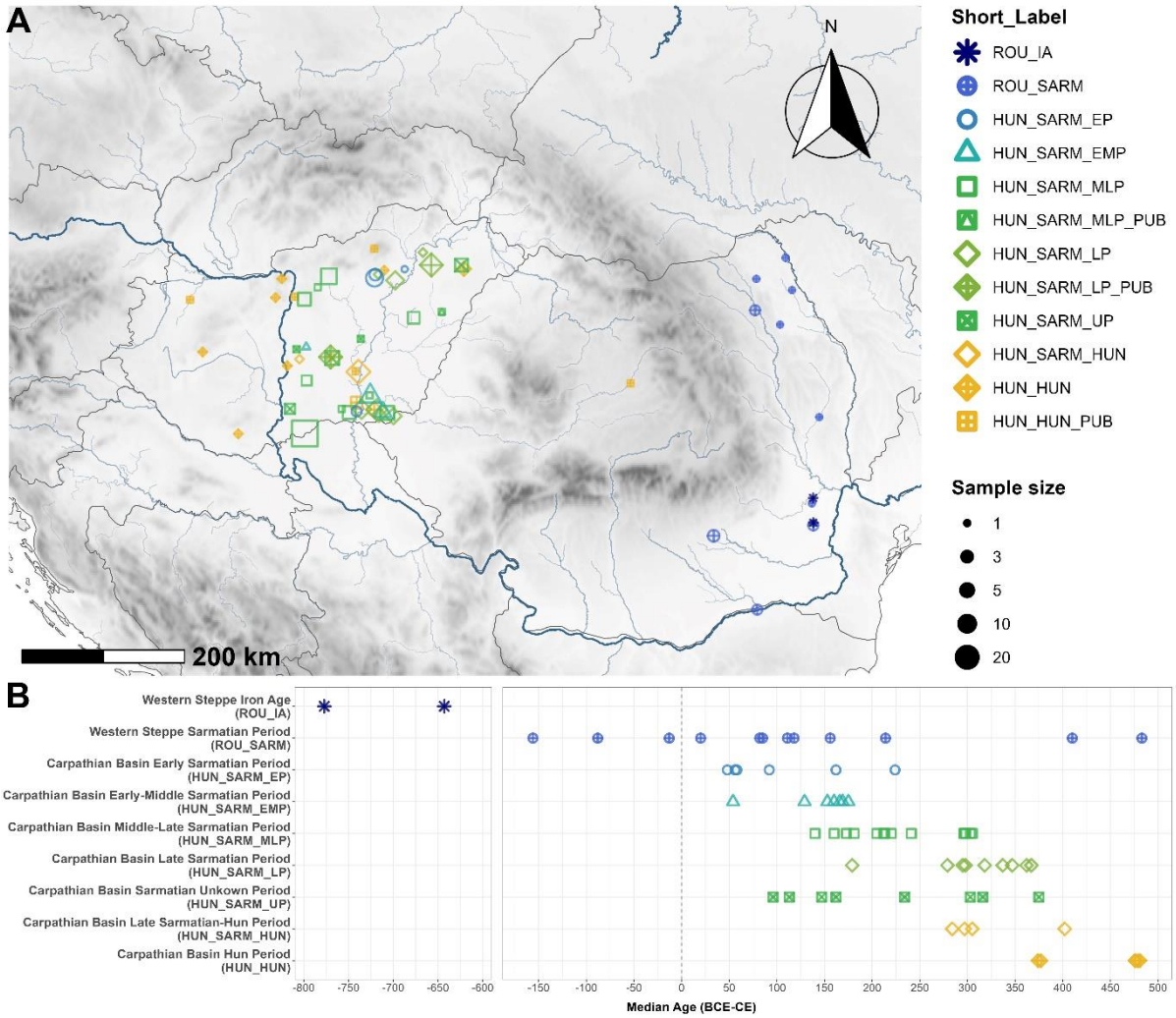


Figure 3: A) Spatial distribution of the samples analysed in this study. B) Temporal distribution of the samples analysed by radiocarbon measurement.

IV.2 Analysing genetic structure

To determine the underlying structure among our samples in a hypothesis independent manner we used PCA and ADMIXTURE analysis (details are described in sections **III.2.1** and **III.2.2**).

IV.2.1 PCA

We calculated primary PC axes from a contemporary Eurasian population set (see **Extended Figure 2** and Maróti et al., 2022) and projected our samples and all available ancient Sarmatian individuals from the AADR dataset onto these axes (**Figure 4**)

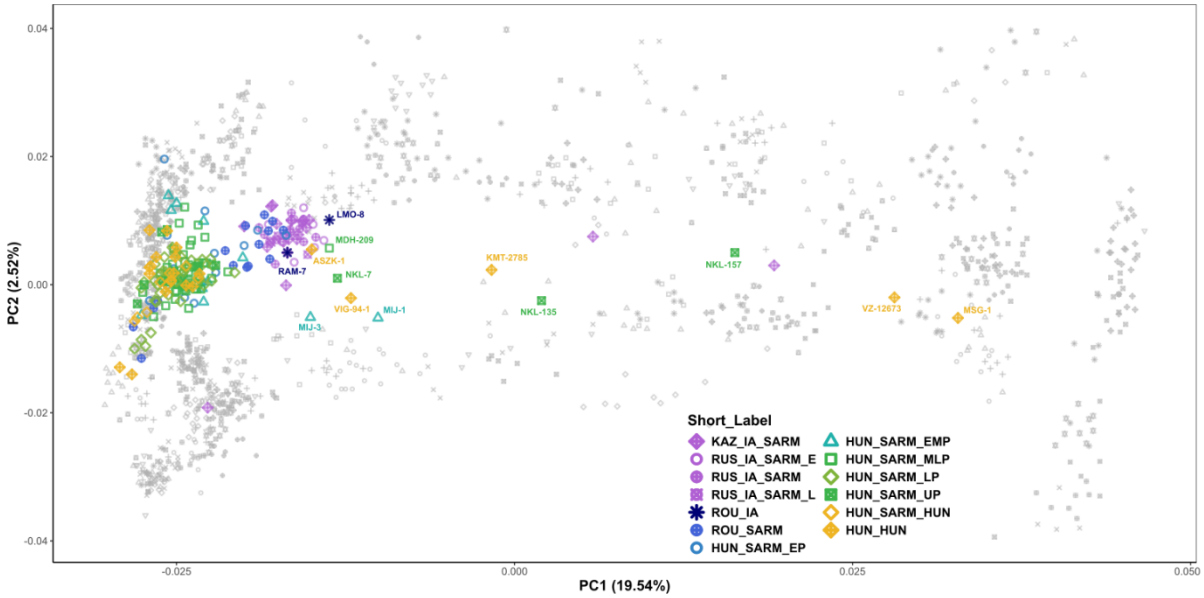


Figure 4: PCA of the studied individuals and all available Sarmatians from the literature projected onto a modern Eurasian background (gray symbols).

Besides some outliers, the Sarmatians of the Russian and Kazakh Steppe (signified with lavender colour on **Figure 4**) form a tight group despite their very wide spatio-temporal distribution. Similarly, our studied Sarmatians (signified with a blue-green palette on **Figure 4**) are also quite homogeneous, with most of the samples mapping closely together. Although historical and archaeological sources emphasize their continuity, the majority of the Sarmatians excavated in the Carpathian Basin occupy a distinct position much closer to the contemporary Central and Southern European populations. However, three groups stand out as exceptions to this pattern. The ROU_IA group and most of the ROU_SARM individuals (filled blue circles on **Figure 4**) overlap with the Steppe Sarmatians, while the HUN_SARM_EP individuals (empty blue circles on **Figure 4**) appear to occupy an intermediate position between these groups and the rest of the Carpathian Basin Sarmatians.

Besides this, the PCA clearly shows the presence of additional genetic outliers with ancestry distinct from the Steppe Sarmatians. Some individuals show strong genetic affinity towards present-day Northern European populations, while others fall more closely to contemporary Southern Europeans. Notably, certain outliers exhibit clear genetic links to Inner Asia, suggesting connections between the studied population and other nomadic groups independent of the Steppe Sarmatians. However, as similar ancestry is also observed – albeit marginally – among some Steppe Sarmatian outliers, it remains possible that these genetic outliers arrived alongside the Steppe Sarmatians.

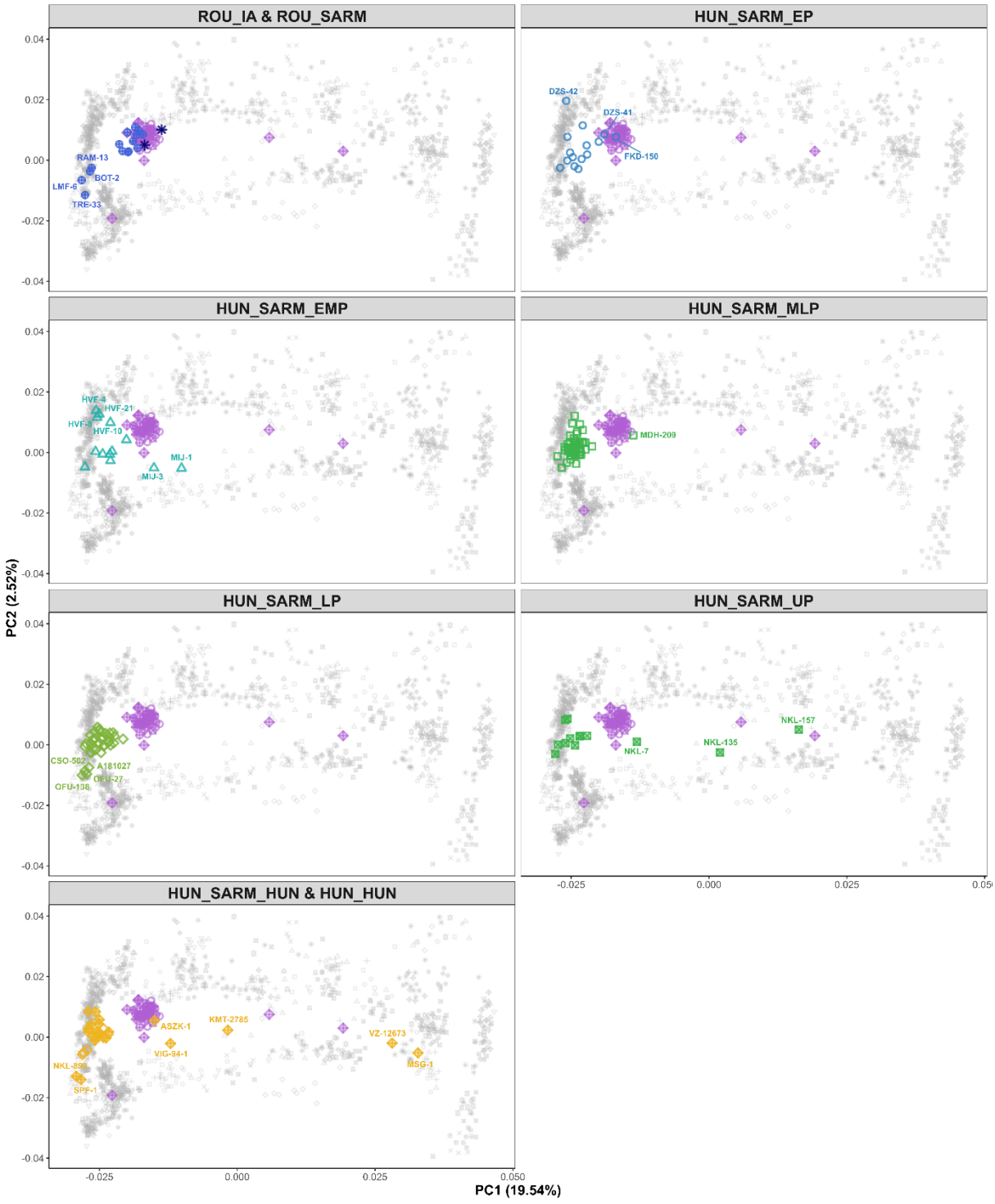


Figure 5: PCA of the studied individuals, with groups from different periods displayed separately. Labels are identical to those in Figure 4. Steppe Sarmatians are shown with lavender shapes in each plot as comparison. Notable outliers are labelled by their name in each subplot.

When examining the different periods separately (**Figure 5**), distinct genetic patterns emerge, diverging over the progressive timeframe. The Sarmatians from Romania (ROU_IA & ROU_SARM on **Figure 5**) form two distinct groups. While the majority of them overlap with

the Steppe Sarmatians, four individuals exhibit a clear genetic affinity toward modern Southern Europeans. The presence of this genetic signature outside the Carpathian range suggests that some Steppe Sarmatians may have already carried a “mixed” genetic background upon their arrival at the foothills of the Carpathian Basin. Additionally, it implies that individuals with Southern European-like ancestry were not only present among the local population but also among incoming groups. This ancestry type becomes increasingly significant in later Sarmatian Periods and plays a particularly prominent role in the subsequent Avar Period (see Maróti et al., 2022). Notably, the two Iron Age individuals from Romania (coloured with navy blue on **Figure 5**) also overlap with the Steppe Sarmatians, despite preceding the appearance of the earliest Sarmatians in this region by at least 300 years.

The Carpathian Basin Sarmatians from the Early Period (HUN_SARM_EP in **Figure 5**) appear to have already undergone substantial genetic change, as many individuals shift toward modern Central Europeans. Nevertheless, at least two individuals (DZS-41 and FKD-150) still align closely with the Steppe Sarmatians, while one individual (DZS-42) exhibits a strong genetic affinity to modern Northern Europeans. This genetic pattern becomes more complex in the Early-Middle Period (HUN_SARM_EMP in **Figure 5**), where we observe an increased presence of individuals with Northern European ancestry (HVF-4, HVF-8, HVF-21), alongside others with a distinct Steppe ancestry seemingly independent of the Steppe Sarmatians (MIJ-1, MIJ-3).

By the Middle-Late and Late Periods (HUN_SARM_MLP and HUN_SARM_LP in **Figure 5**), a clear population consolidation emerges. The previously observed "Central European"-like genetic ancestry becomes more stratified, with the vast majority of individuals clustering tightly together. Outliers become rare (MDH-209), despite these periods encompassing the largest number of studied individuals.

Distinct genetic ancestries reappear in the subsequent Hun Period (HUN_SARM_HUN & HUN_HUN on **Figure 5**), where new outliers emerge with clear genetic affinity to contemporary East Asians (VIG-94-1, KMT-2785, VZ-12673, MSG-1). Nevertheless, the majority of individuals maintain the tight genetic profile established in earlier periods. Samples with uncertain periodization (HUN_SARM_UP in **Figure 5**) also include some outliers (notably NKL-7, NKL-135, and NKL-157). Considering previous observations, these individuals are most likely from the Hun Period.

IV.2.2 ADMIXTURE

ADMIXTURE analysis was prepared using an extensive list of exclusively ancient individuals (**Table E1**). We decided to deviate from our previous analysis (published in Maróti et al., 2022) due to the substantial increase in available ancient genomic data, which allowed for a more comprehensive representation of the ancient Eurasian metapopulation. This approach also enabled us to utilize the 1240K SNP (Fu et al., 2015; Haak et al., 2015; Mathieson et al., 2015) set instead of the more restrictive HO SNP set (Patterson et al., 2012), significantly increasing the number of overlapping informative markers. Cross-validation error calculations indicated that modelling six ancestral components (K=6) produced the most consistent results (**Table E1b**).

Results indicate that we could identify the usual macroregional ancestral components (**Figure 6**) that we termed Western Hunter Gatherer (WHG), Anatolia Neolithic (ANAT_N), Iran Neolithic (IRAN_N), Ancient North Eurasian (ANE) and Ancient North Asian (ANA). Additionally, a peculiar component appeared which was most pronounced in our newly sequenced Carpathian Basin individuals (thus termed CB, blue in **Figure 6**). This seemed to represent a stable mixture of WHG, IRAN_N and ANE that was retained in the Carpathian Basin population through the ages. The appearance of this stable genomic composition may have been caused by the overrepresentation of this region with many recently published samples that we included (from Gnecchi-Ruscone et al., 2022; Maróti et al., 2022; Patterson et al., 2022).

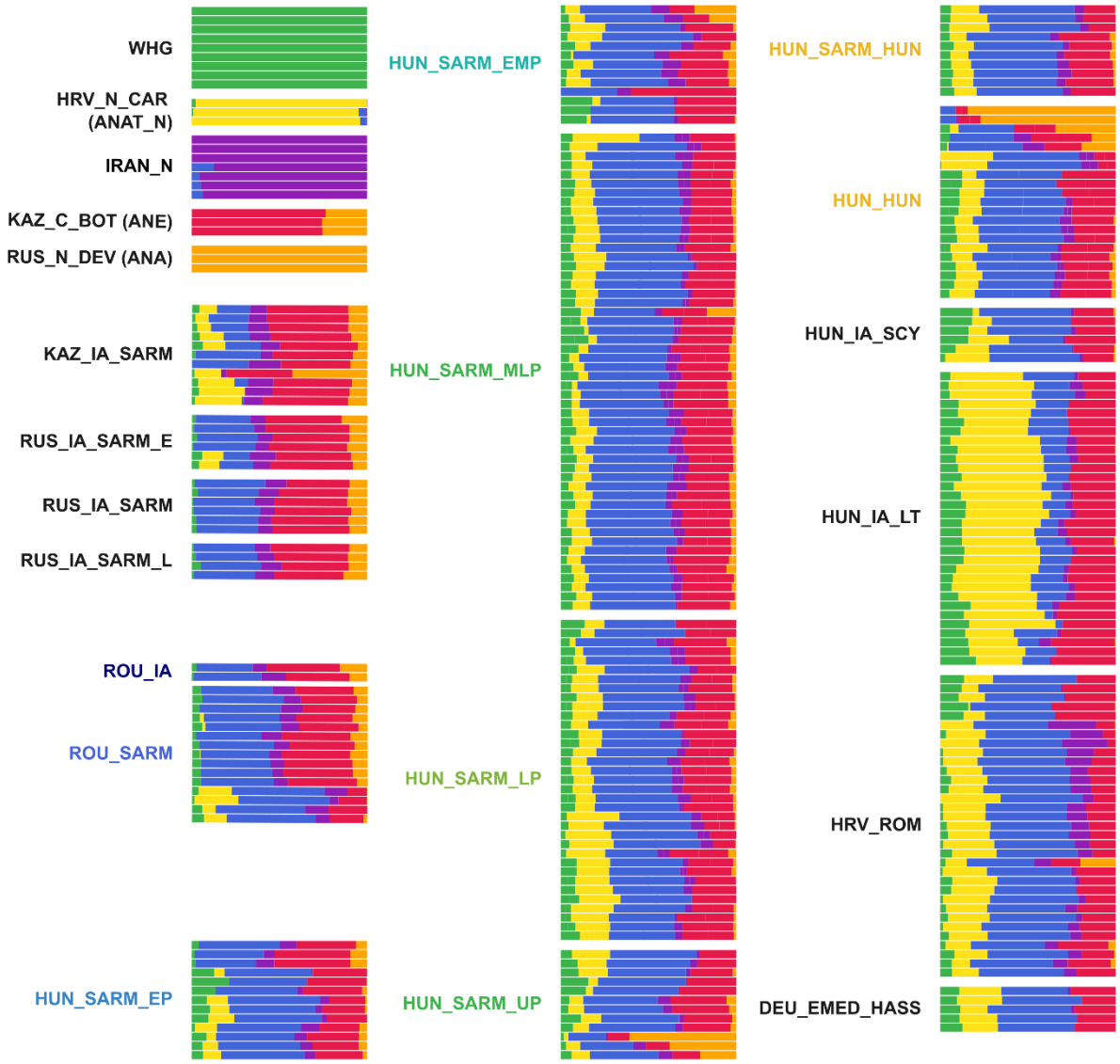


Figure 6: Unsupervised ADMIXTURE analysis results (K=6) of the studied individuals together with the available Steppe Sarmatians (for the Short Label abbreviations and the rest of the results see **Table E1**). Groups best representing the main ancestral components are shown at the top left side. On the right we also show selected contemporary populations from the Carpathian Basin for comparison.

Most Sarmatian and Hun-period individuals from the Carpathian Basin closely resemble contemporaneous populations from the region (Croatia Roman period - HRV_ROM on **Figure 6**) or populations from the immediately preceding period (Hungary Scythians - HUN_IA_SCY and Hungary Celtic - HUN_IA_LT on **Figure 6**). However, a distinctive feature of our studied groups is the presence of a small but substantial fraction of the ANA-related component (orange in **Figure 6**). This component is highest in the earliest groups (HUN_SARM_EP and EMP on **Figure 6**) and appears to decline over the progressive periods. The ANA component is much

more pronounced in the Steppe Sarmatians and Romanian Sarmatians, who have very similar genome compositions, corresponding to their overlap on the PCA and their significant shift from European groups. The PCA Asian outlier individuals also stand out in the ADMIXTURE analysis due to their substantial ANA component, which is much larger than that of the Steppe Sarmatians (like MIJ-1 and MIJ-3 from the HUN_SARM_EMP group or VZ-12673 and MSG-1 from the HUN_HUN group, for further examples compare **Figures 5** and **6**).

The two Iron Age individuals from Romania show very surprising ADMIXTURE patterns, displaying identical component ratios with the RUS_IA_SARM group (ROU_IA on **Figure 6**) which is also reflected by their close PCA positions (**Figure 5**). This sharply distinguishes them from the contemporaneous Scythian and Celtic populations (HUN_IA_SCY or HUN_IA_LT on **Figure 6**) that were historically known to inhabit this area. A small exception to this are the handful of published individuals attributed to the Cimmerian culture (Järve et al., 2019; Krzewińska et al., 2018). They display substantial East Asian genetic affinity, however they do not seem to have exerted a strong genetic impact on their surroundings, as this East Asian genetic influence only reappears much later, mediated by the Sarmatians. Interestingly, the two available "Cimmerian" individuals with sufficient genome coverage (cim357, cim359 from Krzewińska et al., 2018) show very similar component ratios to those observed in our ROU_IA individuals (**Table E1a**).

IV.3 Reanalysing the Steppe Sarmatians

Since historical sources suggest a direct link between the Sarmatians of the Carpathian Basin and those from the central steppes, we aimed to reanalyse the published Steppe Sarmatian data in order to create homogeneous reference sets for further analyses. As discussed in **I.2.3**, despite their vast spatial distribution and significant cultural influence (**Figure 7**), only a limited number of Steppe Sarmatian genomes are available in the literature. Moreover, these genomes span a very broad temporal range, and the lack of detailed contextual background further complicates drawing reliable conclusions regarding their genetic makeup.

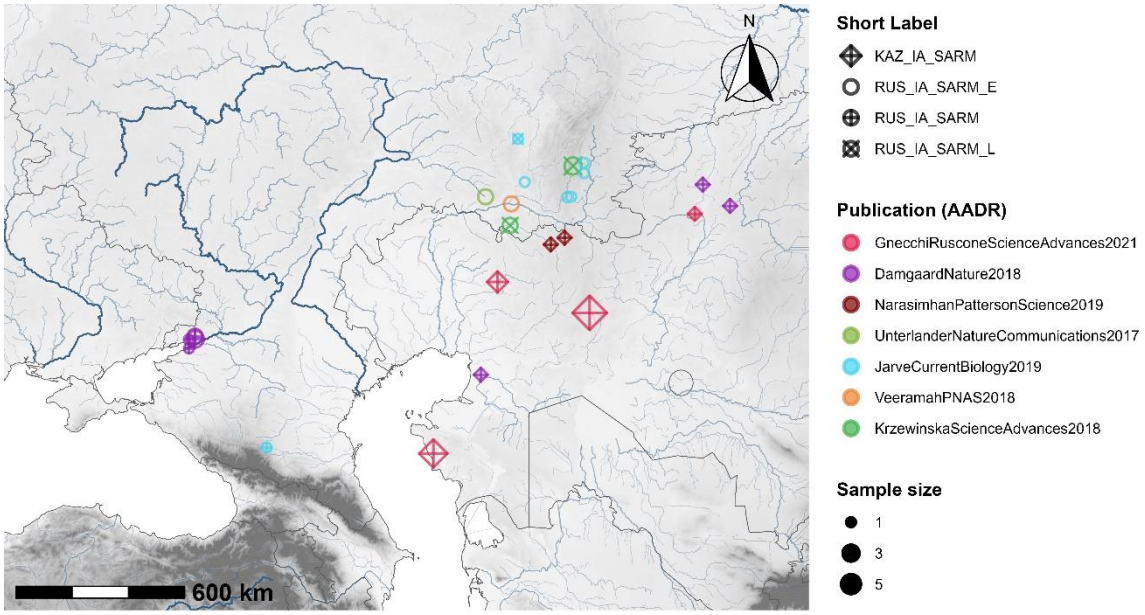


Figure 7: Geographical distribution of the Steppe Sarmatian samples. Point colours represent the parent articles. A minor jitter effect was added to show off cemeteries with very close geographical locations. The point sizes represent sample size.

Nevertheless, we ran PCA, ADMIXTURE and qpAdm analyses to detect any previously unnoticed subtle differences in the published genomes (**Figure 8**). While the Sarmatian individuals from the steppe regions show strong homogeneity on the PCA (**Figure 8A**), that persists across time, the ADMIXTURE analysis identified at least two marginally distinct clusters (**Figure 8B**). These findings were also supported by qpAdm analyses (data not shown).

We finally constructed two genetically homogeneous groups formally titled STEPPE_IA_SARM_URAL (including chy001, tem001, tem002 from Krzewińska et al. (2018); DA134, DA139, DA141, DA143, DA144, DA145, DA26, DA30 from Damgaard et al. (2018); and MJ-41, MJ-44 from Järve et al. (2019)) and STEPPE_IA_SARM_STEPPE

(including AIG001, AIG002, AIG006, BSB001, BSB003, CLK001, KBU001, KBU002, KSK002 from Gnecchi-Ruscone et al., 2021; I11537, I11540 from Unterländer et al., 2017; MJ-39 from Järve et al., 2019; and Pr4 from Veeramah et al., 2018). These groups were used as source populations in subsequent qpAdm modelling frameworks.

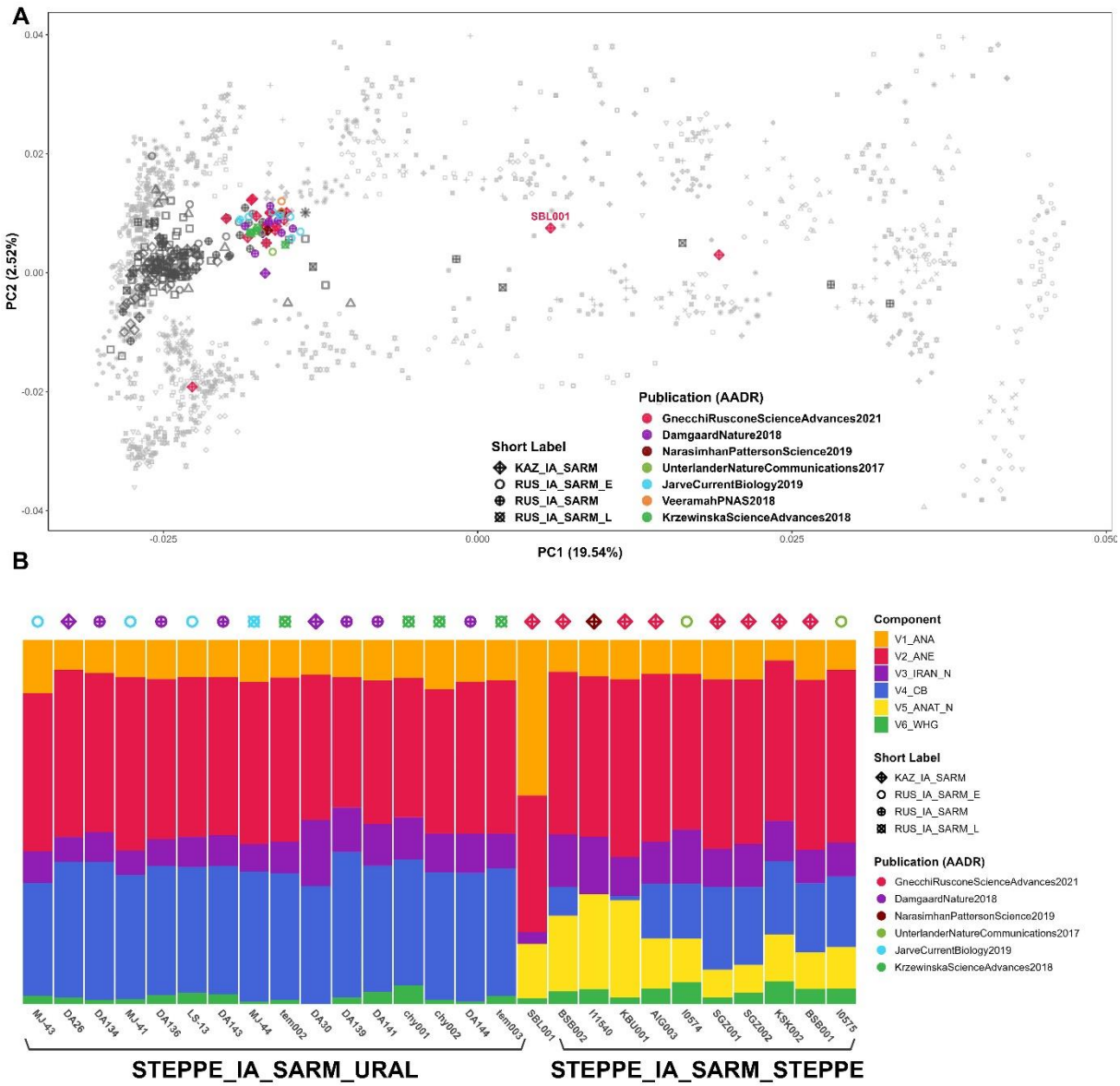


Figure 8: PCA and ADMIXTURE results of the Steppe Sarmatians. A) PCA analysis of the available Steppe Sarmatian genomes projected onto the same modern Eurasian population set as in **Figure 2**. Points are coloured according to the parent publication. B) ADMIXTURE results of the Sarmatian samples with appropriate genome coverage distinguishing two groups.

The PCA reveals several outlier individuals who are shifted toward Asia (**Figure 8A**). Unfortunately, most of these had coverage too low for ADMIXTURE and further analyses, except for one individual, SBL001 (**Figure 8B**). Nevertheless, it is remarkable that, despite the

broad region and timeframe considered, the overwhelming majority of the individuals map within a highly homogeneous genetic group. Moreover, the presence of outliers with presumably elevated Inner Asian ancestry aligns with our findings of similar outliers in the HUN_SARM_EMP and later groups. This suggests the possibility that distinct steppe groups may have accompanied the Sarmatian migrants into the Carpathian Basin.

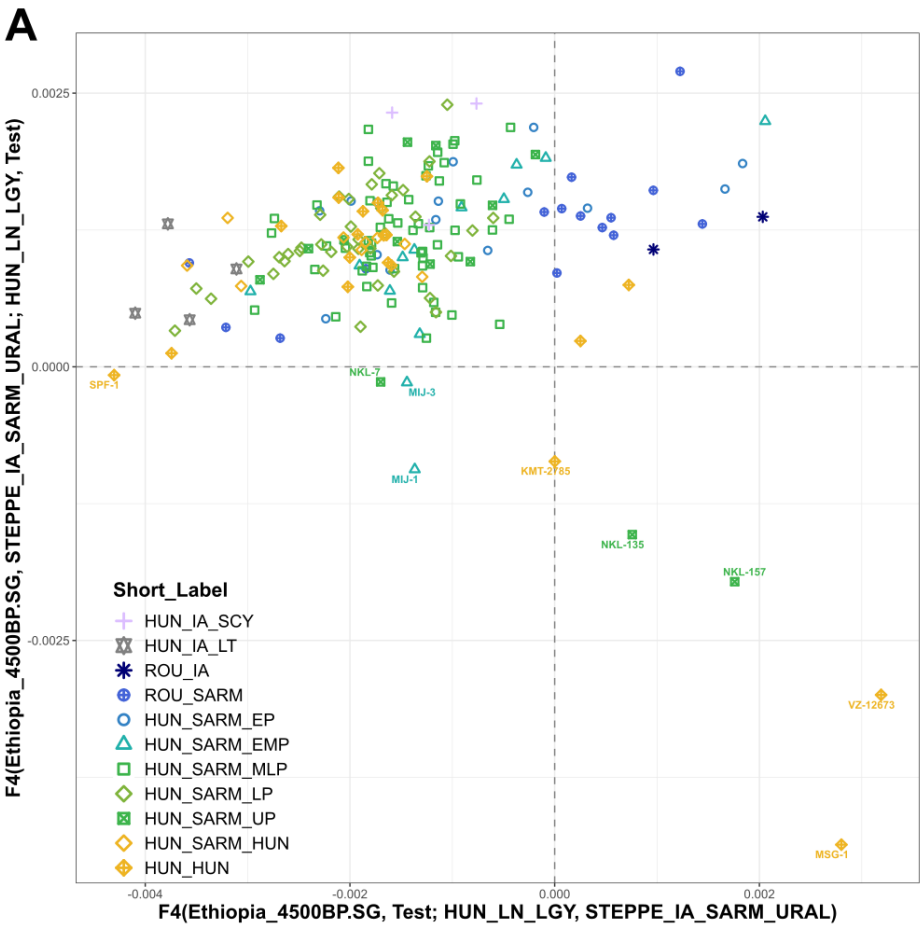
IV.4 F-statistics

While population structure analysis shows our samples to be clearly distinct from the Steppe Sarmatians, they also display some similarities – namely their elevated East Asian ancestry – that separates them from their contemporary neighbours. Furthermore, historical and archaeological sources indicate a clear connection between these groups (Istvánovits & Kulcsár, 2017). Therefore, we aimed to explore their deeper genetic affinity with the Steppe Sarmatians.

IV.4.1 F4-statistics

We applied F4-statistics, in which we co-analysed the Sarmatian and Hun-period samples with populations that most likely exemplify the immediately preceding inhabitants of the Carpathian Basin (for the obtained F4 values see Schütz et al., 2025). First, we measured the direct affinity of the samples towards the Steppe Sarmatians against a Late Neolithic Carpathian Basin population (Lengyel culture) possibly representing local elements, with the statistics: **F4(Ethiopia_4500BP, Test; HUN_LN_LGY, STEPPE_IA_SARM_URAL)**. Positive values in this statistic indicate a major affinity towards the Steppe Sarmatian proxy, while negative values show more shared drift with the local proxy. This was plotted together with another combination: **F4(Ethiopia_4500BP, STEPPE_IA_SARM_URAL; HUN_LN_LGY, Test)**, which uses the same references, but actually measures the samples' affinity towards the Sarmatian proxy while excluding their shared drift with the local proxy (**Figure 9**, for further discussion on this property of the F4-statistics see **III.2.3**).

Figure 9A shows that most of the samples (including the two other Iron Age Carpathian Basin groups, Hungarian Scythians – HUN_IA_SCY and Celtic individuals – HUN_IA_LT) shares the majority of their markers with the local proxy, while only a handful of individuals (especially from the ROU_SARM and HUN_SARM_EP groups) fall on the positive side of the X axis. This is further supplemented by the Y axis where almost all individuals map on the positive side displaying at least a limited affinity toward the Steppe Sarmatian proxy in addition to their local affinity. The genomes located in the upper right quadrant exhibit strong Steppe Sarmatian affinity. Notably, these samples also display overlapping PCA positions and similar ADMIXTURE patterns to those of the Steppe Sarmatians (see **Figures 4, 5** and **6**).


A
F4(Ethiopia_4500BP.SG, STEPPE_IA_SARM_URAL; HUN_LN_LGY, Test)
F4(Ethiopia_4500BP.SG, Test; HUN_LN_LGY, STEPPE_IA_SARM_URAL)
0.0025
0.0000
-0.0025
-0.004
-0.002
0.000
0.002
SPF-1
NKL-7
MIJ-3
MIJ-1
KMT-2785
NKL-135
NKL-157
VZ-12673
MSG-1
Short_Label
HUN_IA_SCY
HUN_IA_LT
ROU_IA
ROU_SARM
HUN_SARM_EP
HUN_SARM_EMP
HUN_SARM_MLP
HUN_SARM_LP
HUN_SARM_UP
HUN_SARM_HUN
HUN_HUN

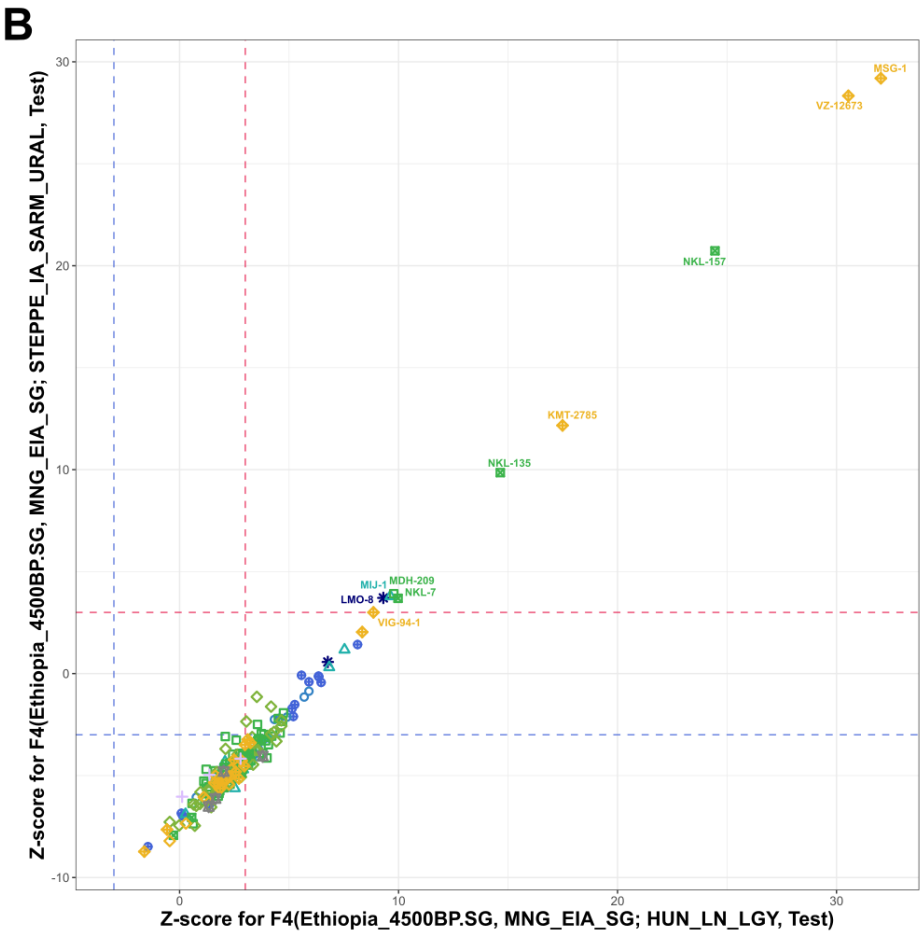

B
Z-score for F4(Ethiopia_4500BP.SG, MNG_EIA_SG; STEPPE_IA_SARM_URAL, Test)
Z-score for F4(Ethiopia_4500BP.SG, MNG_EIA_SG; HUN_LN_LGY, Test)
30
20
10
0
-10
0
10
20
30
MSG-1
VZ-12673
NKL-157
KMT-2785
NKL-135
MIJ-1
MDH-209
NKL-7
LMO-8
VIG-94-1

Figure 9: A) F-scores of Steppe Sarmatian affinity measured in our samples and other Iron Age groups against local Carpathian Basin ancestry. The X axis shows the two-way affinity of the samples to either the local or the Sarmatian proxy, while the Y axis measures the affinity toward the Steppe Sarmatian proxy after excluding their local marker set. B) Z-scores of East Asian affinity of the samples measured by excluding their local marker set (X axis) or their marker set potentially shared with the Steppe Sarmatians (Y axis). Blue segmented lines indicate the negative significance threshold (-3), red lines indicate the positive significance threshold (3). Samples with exceptionally high East Asian affinity are labelled.

A few individuals in the lower right quadrant of **Figure 9A** show strong negative F-scores on the Y axis but retain Steppe Sarmatian affinity on the X axis. They also have an elevated East Asian genomic component, as seen in PCA and ADMIXTURE, distinguishing them from the Steppe Sarmatian proxy. On the other hand, these samples share significant drift with Steppe Sarmatians on the X axis, likely due to their East Asian ancestry. This raises the possibility that other samples with Steppe Sarmatian affinity on the Y axis might show similar patterns because of their elevated East Asian ancestry.

To address this uncertainty, we conducted two additional F4 analyses: 1.) F4(Ethiopia_4500BP, MNG_EIA_SG; HUN_LN_LGY, Test) that measures the East Asian affinity of the samples using Mongolia_EIA_SlabGrave as proxy, while excluding markers shared with the local proxy. 2.) F4(Ethiopia_4500BP, MNG_EIA_SG; STEPPE_IA_SARM_URAL, Test) that assesses the same East Asian affinity but excludes markers shared with the Steppe Sarmatian proxy. In **Figure 9B**, we plotted Z-scores instead of F-values, as the significance level of the statistics provides a more precise answer to our question.

The X axis of **Figure 9B** demonstrates that all individuals on the right side of **Figure 9A** show significant shared drift with the East Asian proxy beyond their local ancestry. However, the Y axis reveals that the East Eurasian affinity of the samples in the upper right quadrant of **Figure 9A** is equivalent to the Uralic Sarmatians, as this affinity falls around zero when only considering markers not shared with the Uralic Sarmatian. In contrast, individuals in the lower right quadrant of **Figure 9A** show significant Z-scores on both axes, indicating a higher level of East Asian genetic affinity that cannot be attributed to the Steppe Sarmatian proxy alone. The presence of these individuals suggests the existence of a source distinct from the Sarmatians.

While the two Iron Age groups from the Carpathian Basin, HUN_IA_SCY and HUN_IA_LT, show some affinity toward the Steppe Sarmatian proxy in **Figure 9A** (especially

HUN_IA_SCY), this affinity likely stems from a different component. This is supported by **Figure 9B**, where these groups fall well below any significance line, indicating no detectable East Asian genetic affinity. Notably, our two Iron Age individuals from Romania (ROU_IA) display strong affinity towards the Steppe Sarmatians on both **Figures 9A** and **B**, further supporting the association of these two groups.

IV.4.2 qpAdm

The results of the F4-statistic analyses were also verified using the hypothesis testing algorithm qpAdm (**Figure 10** and **Table E2**). As our main goal was to determine whether Steppe Sarmatians are essential for modelling the Carpathian Basin Sarmatians, we devised a purposeful qpAdm analysis framework to this end.

Based on preliminary qpAdm runs, we assembled a comprehensive set of 15 source populations to best represent the potential local Carpathian Basin elements in most individuals. The previously assembled Steppe Sarmatian groups (described in detail in **IV.3**) were used as the representative sources for our proposed Sarmatian immigrants. Additionally, we included a further 8 sources comprised of other Central and Inner Asian populations that had known connections to the Carpathian Basin or could represent further Asian immigrants independent of the Sarmatians. For further details on the analysis framework and exact composition of the LEFT and RIGHT populations see **III.2.4** and **Table E2d**.

In our initial analysis, we successfully obtained appropriate 2 or 3-source models for all individuals using the local European and Steppe Sarmatian sources except for 12 samples. Next we slightly modified our analysis framework by including further source candidates from a wider spatio-temporal range and successfully obtained valid models for all of the remaining outliers except for one individual (HVF-10).

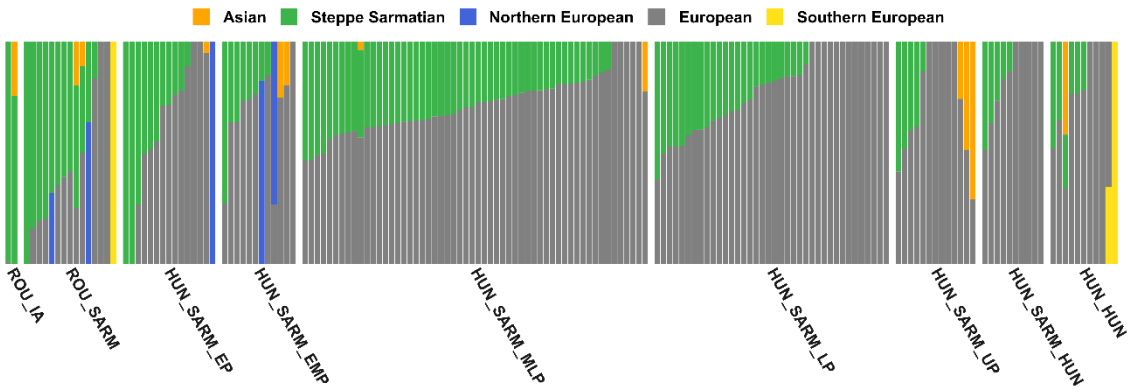


Figure 10: Simplified summary of the qpAdm results based on **Table E2**. Models are arranged according to similarity of components. For the interpretation of the colour scheme see **III.2.4**.

IV.4.2.1 Genetic composition of the Sarmatians from the Carpathian Basin

Two individuals, DZS-41 and FKD-150, who were among the earliest Sarmatians in the Carpathian Basin according to archaeological data and radiocarbon dating, along with one Romanian Sarmatian (OSU-1), produced unambiguous models, in which they formed a genetic clade with the Steppe Sarmatians (the first column of ROU_SARM and first two columns of HUN_SARM_EP on **Figure 10**). Many of the samples however, exhibited a pattern similar to that observed in the F4 analysis, displaying a marginal yet detectable East Asian affinity (**Table E2a-b**). The exact source of this was unfortunately hard to distinguish due to its low contribution, even by our model competition approach. Almost half of the obtained feasible models fell in the category of this uncertain composition, where usually 70-95% of the composition could be modelled from one of our "local" sources with either 20-30% Steppe Sarmatians or 5-15% East Asian/Central Asian sources. This ambiguity raises questions regarding the precise origin of the East Asian component. Fortunately, we also obtained several well-defined models that provided clearer insights.

On one hand, we identified multiple feasible models that unambiguously indicate Steppe Sarmatian ancestry as a minor component (like MDH-265 or A181015). In many cases these individuals appeared in the same cemetery as others with uncertain models. Combined with the results from the Early Period Sarmatians, who exhibit a cladal relationship with the Steppe Sarmatians, these findings strongly suggest that Steppe Sarmatians were the most likely source of this ancestry.

On the other hand, we also identified individuals whose models clearly indicated Central or East Asian ancestry (e.g. MIJ-1, MIJ-3, MDH-209 or NKL-135). These were the same individuals that separated on PCA (**Figures 4** and **5**) and exhibited significant East Asian affinity in the F4-statistic analysis (**Figure 9**). Furthermore, they never displayed ambiguous models, suggesting that when the East Asian component was substantial enough to be detected by simple F4 analysis, its source can be clearly distinguished by the qpAdm analysis.

Taken together, these findings support the assumption that cases of uncertain ancestry are more likely to contain Steppe Sarmatian admixture. This ancestry was empirically more widespread among the samples and present from the very beginning of the Sarmatian Period. In contrast, "Sarmatian-independent" East Asian ancestry was only detected in a few individuals through

multiple analyses and always unambiguously. Thus, the hypothesis tests confirms that most of our studied samples exhibit at least some shared drift with the Steppe Sarmatians supporting the claims of the historical sources. The seemingly distinct PCA distribution of the Carpathian Basin Sarmatians (**Figure 4**) likely reflect a genetic cline between the Steppe Sarmatians and a still underrepresented local population, though our samples occupy a highly stratified position with only minor deviations from the population average.

IV.4.2.2 Ancestry composition before and after the Sarmatian Period of the Carpathian Basin

The Sarmatians from Romania show a much stronger affinity towards the Steppe Sarmatians attested both by their PCA position and qpAdm models (ROU_SARM on **Figures 4, 5** and **10**, also see **Table E2a-b**). With the exception of four individuals (BOT-2, LMF-6, RAM-13, and TRE-33), the majority of the ROU_SARM group (73%) displayed predominant Steppe Sarmatian ancestry. Among the exceptions, LMF-6 was a Southern European outlier (**Table E2c**), TRE-33 could be modelled entirely from “local” sources, while BOT-2 and RAM-13 exhibited the previously discussed uncertain ancestry pattern. This strong genetic connection supports the historically suggested Lower Danube migration route of the Sarmatians into the Carpathian Basin. However, we have to note that sampling was not undertaken from the northern vicinity of the Carpathian range, thus definitive conclusions should by avoided until further investigations.

Furthermore, a problem arises: nearly all individuals in this group also carry at least a minor genetic component from one of our “local” sources, indicating an actively transitioning population. This suggests that the dataset used to represent the "locals" does not strictly correspond to the geographic borders of the Carpathian Basin but rather reflects a broader metapopulation, with which the migrating Sarmatians may have already interacted on the western steppe.

The distribution of non-local ancestry components appears to shift across time. Most individuals with major Steppe Sarmatian ancestry are found in the HUN_SARM_EP and HUN_SARM_MLP groups, while only one such individual (ASZK-1) is present after the HUN_SARM_LP phase (and even this sample carries an additional East Asian minor component). In contrast, the individuals with the highest proportion of East Asian ancestry (KMT-2785, MSG-1, and VZ-12673) are concentrated in the HUN_HUN group, signalling potential population changes (the qpAdm models of the reanalysed samples from Maróti et al. (2022) are present in that article).

While these patterns are clear, it is important to note that most individuals from the HUN_SARM_LP, HUN_SARM_HUN and HUN_HUN groups still retain at least a marginal Steppe Sarmatian ancestry, connecting them to the earlier periods. Interestingly, nearly half of the unambiguously "local" individuals also belong to the HUN_SARM_LP or later groups, further emphasizing the complex demographic dynamics of the region.

IV.4.2.3 Modelling the qpAdm "outliers"

We identified 12 individuals (here referred to as "outliers"), who could not be successfully modelled within our initial analysis framework. In our previous article (Maróti et al., 2022), we encountered similar cases where certain individuals could only be modelled by incorporating additional sources that initially seemed unlikely to have had a significant genetic impact on the region. Drawing from that experience, we refined our framework by introducing additional source populations, allowing us to obtain feasible models for all but one of the "outliers" (**Table E2c**).

The sole unmodelled sample, HVF-10, showed very high F3 values with Mongolia_Chalcolithic_Afanasievo_1 (F3 = 0.288) and other Early Bronze Age groups, while displaying only 2 ancestral components in the ADMXITURE analysis. This suggests the possibility of a misdated Bronze Age individual. However, HVF-10 shares reasonably long IBD segments with other individuals from the Sarmatian and subsequent periods of the Carpathian Basin, indicating that they are probably contemporaneous with others from the same cemetery.

A portion of the outlier individuals exhibited confirmed signs of affinity towards modern day Northern European populations, as they could only be modelled fully or partially using Scandinavian-related sources (e.g., Germany_Hassleben_Germanic_elite_2 or Estonia_BA), which were clearly distinct from any ancient populations in the immediate vicinity of the Carpathian Basin. This suggests the possibility of long-range migrations, potentially linked to the Roman Empire's influence.

In our previous study, we found that many individuals, particularly from the Avar Period, required Neolithic sources (e.g., Hungary_MN_LBK or Hungary_Tisza_LN) for successful modelling. This was much more limited here, as only three individuals required similar sources (LMF-6, NKL-899 and SPF-1). This, along with the PCA results, hints at the possibility that the subpopulation observed in the Avar Period may have arrived in the Carpathian Basin after the Sarmatian Period.

Five individuals (RAM-7, LMF-7, POG-10, TAF-11, FKD-150) showed strong genetic affinity to the core Steppe Sarmatians, as indicated by their PCA positions (**Figures 4** and **5**), yet our original Steppe Sarmatian reference groups were insufficient to model them accurately. By replacing these reference groups with constituent individual samples from the Steppe Sarmatian group, as well as with selected Early Period Sarmatians from the Carpathian Basin (e.g., DZS-41), we obtained feasible models for all five individuals. Interestingly, one of the ROU_IA individuals (RAM-7) was also among these samples and could be feasibly modelled from one of our HUN_SARM_EP individuals (DZS-41). This finding reinforces the strong genetic affinity between these Iron Age individuals and the Steppe Sarmatians excavated much later and from distant locations.

IV.5 IBD analysis

In order to explore the genealogical links across different geographic regions and time periods in Central Europe and the Central Steppe, we conducted Identity-by-Descent (IBD) analyses. For this purpose, we selected and imputed 504 individuals spanning from the Iron Age to the early Middle Ages additionally to the 158 genomes presented in this study.

The main criteria for selecting individuals were their spatio-temporal origin and the library preparation method, excluding capture-enriched genomes to avoid erroneous genotype inferences during imputation. We applied a stringent entry threshold for imputation recommended by Rubinacci et al. (2021), requiring a minimum of 0.5-fold coverage and low contamination, resulting in the exclusion of 7 new samples. Additionally, we excluded the 17 Sarmatian individuals published in Gnecchi-Ruscone et al. (2022) due to their capture sequencing method.

We identified IBD genomic segments of at least 8 centimorgans (cM) in length using the ancIBD software (Ringbauer et al., 2024) with optimizations described in **III.2.6**. IBD connection networks were visualized as graphs using the Fruchterman-Reingold (FR) weight-directed algorithm (Fruchterman & Reingold, 1991). In these graphs, individual samples are represented as points (vertices), and IBD connections between points are shown as segments (edges). The algorithm operates iteratively, calculating attractive forces for vertices connected by edges and repulsive forces for points not connected by edges in each cycle. We used the total length of IBDs shared between individuals as weights for calculating the attractive force, making the distance between connected vertices roughly proportional to their genealogical distance.

IV.5.1 IBD sharing network of selected Eurasian groups

First, we investigated the genealogical links between different populations across different time periods. For this reason, we grouped the samples by archaeological period and culture. This allowed us to examine intergroup connections among various European and Carpathian Basin populations, including relevant groups from the Central Steppe and Asia. In these analyses the clouds of groups were handled as single points and a group relation layout was calculated with the FR algorithm, where the weights were the number of IBD connections between each group. This way, the distribution of the clouds themselves actually reflects the connectedness between the groups (**Figure 11A**).

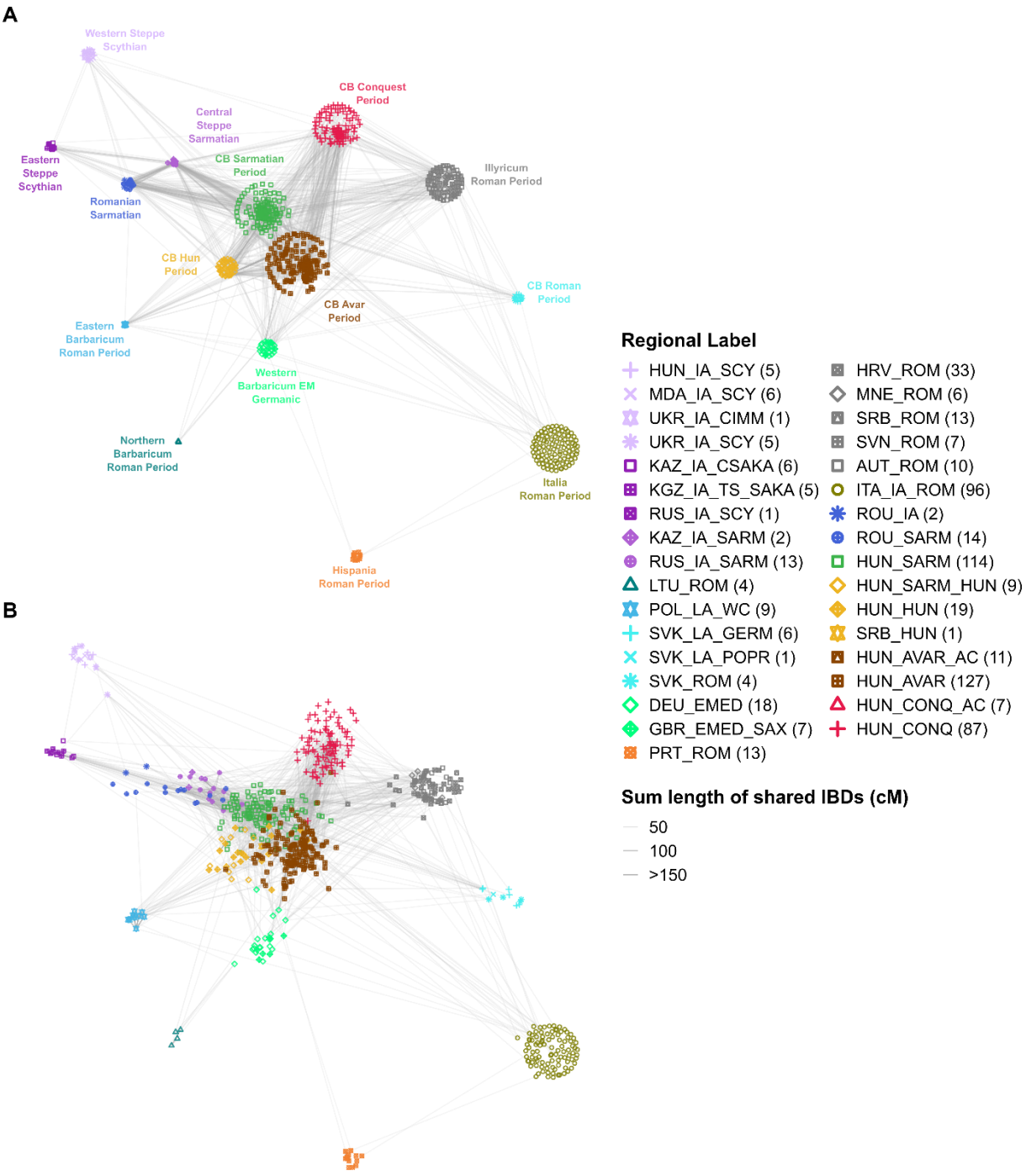


Figure 11: IBD sharing of groups potentially related to Sarmatians A) Intergroup IBD sharing graph of 662 ancient shotgun genomes including the samples presented in this publication. Individuals were grouped according to geographical region and archaeological period. B) Vertices were allowed to reposition driven by the FR weight directed algorithm for 100 iterations. The maximum displacement of the points was reciprocally toned down by the number of outgoing connections (edges pointing outside the respective groups). Only edges representing intergroup connections are plotted. The details on the Regional Label abbreviations can be found in **Table E3b**.

The Sarmatians from the Carpathian Basin (coloured by green on **Figure 11A**) occupy a central position in this plot together with the Hun Period and Avar-period individuals (coloured by gold and brown respectively on **Figure 11A**) published in Maróti et al. (2022). This is not surprising, as these groups are populous with a central spatio-temporal location, hence they have the most opportunities to produce detectable genealogies that connect earlier and later populations. However, this does not undermine the importance of their numerous connections with seemingly distant groups and especially with each other.

Interestingly, steppe-related groups - Scythians, Steppe Sarmatians, and Romanian Sarmatians (coloured by shades of purple, lavender and blue on **Figure 11A**). - all cluster near the Carpathian Basin Sarmatians. These groups share the most IBD segments with the Sarmatians of the Carpathian Basin and with each other. It is especially notable that within these groups, the Western Scythians (HUN_IA_SCY, MDA_IA_SCY, UKR_IA_CIMM, UKR_IA_SCY) exhibit the fewest connections to any other group, including the nearby and contemporary Carpathian Basin Sarmatians, who have significantly more connections with the geographically more distant Steppe Sarmatians.

Also interesting is the Hun-period group, which barely shows any pattern of intragroup relatedness and instead has a disproportionately high ratio of intergroup connections, especially with individuals from the preceding Sarmatian Period and the subsequent Avar Period. This suggests that the Hun-period group may not represent a distinctly separate population.

These observations are further illustrated on **Figure 11B**. In this graph the original positions were given the same way as before, but we ran another FR algorithm with restricted runtime (100 iterations) and let the points displace along their main attraction forces. As can be seen, the Roman provincial samples along the border of the plot barely moved, reflecting their sparse connectedness to the Carpathian Basin individuals. Contrary to this, the Sarmatians excavated in Romania and on the Central Steppe moved rapidly inward, clustering with the Sarmatians of the Carpathian Basin. As expected, the Hun-period individuals also rapidly lost their group coherence and shifted towards the Sarmatian and Avar-period individuals.

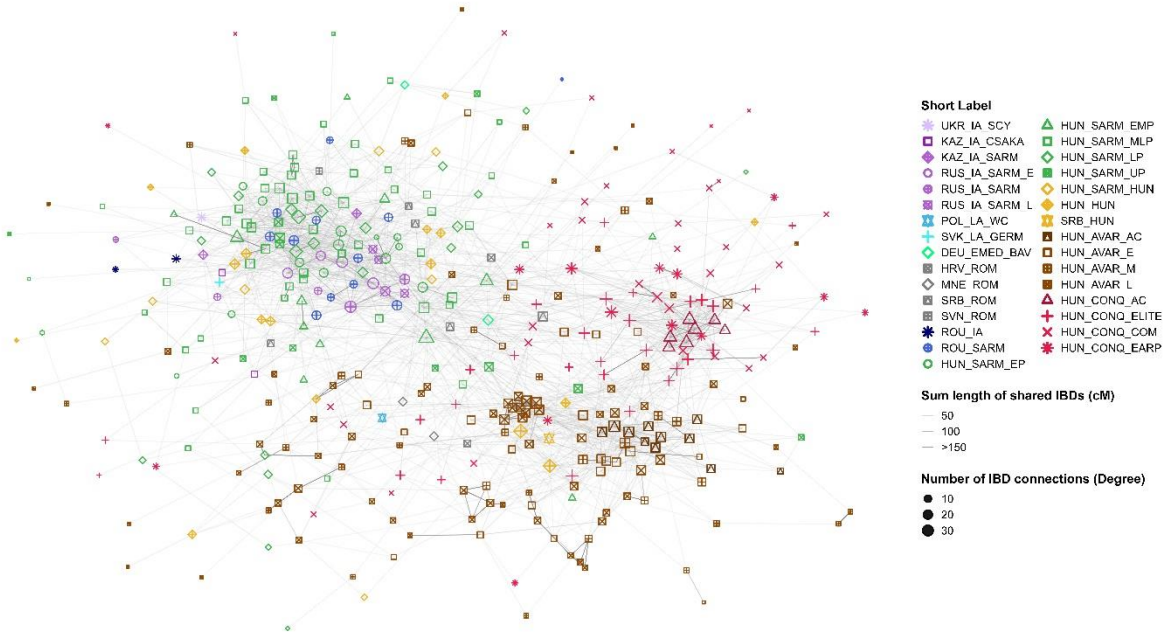


Figure 12: IBD sharing graph of 423 selected individuals. Points represent individual samples. The size of the points is proportional to the number of connections (degree) they have. Edges are shaded according to the total lengths of IBDs shared along them in cM. The details on the Regional Label abbreviations can be found in **Table E3b**.

To better illustrate the IBD sharing patterns on an individual level, we prepared a graph with no grouping (**Figure 12**). Here we included all studied groups from the Carpathian Basin (HUN_SARM, HUN_SARM_HUN, HUN_HUN, HUN_AVAR, HUN_CONQ) and its vicinity (ROU_IA, ROU_SARM) as well as Steppe Sarmatians (KAZ_IA_SARM, RUS_IA_SARM). From the other groups shown in **Figure 11A**, we included only 17 samples that had at least 5 IBD connections to the studied individuals. This threshold was chosen to reduce the number of additional individuals to ≤10%, thereby reducing the clutter of the figure and improving its clarity. This resulted in a collection of 423 individuals, who were plotted with the same original coordinate positions as seen in **Figure 11A**, but the FR algorithm was allowed to freely reposition the points for 1000 iterations.

On **Figure 12**, the three main groups, Sarmatian, Avar and Conquest-period samples from the Carpathian Basin, form distinct clusters, reflecting high intra-period connectivity. The first-generation immigrant "core" individuals of each medieval group appear to occupy central positions, particularly during the Conquest Period (coloured with red on **Figure 12**), where they exhibit very high levels of both intra- and intergroup connectedness. During the Sarmatian Period, this central position is seemingly delegated to the Steppe Sarmatians, within the cluster

of ROU_SARM and HUN_SARM_EP individuals. Conversely, the Hun-period individuals do not form an isolated group but instead seamlessly blend into the Sarmatian and Avar "clouds". Notably, some HUN_SARM_UP (e.g., NKL-157) and HUN_HUN individuals (e.g., MSG-1, VZ-12673, KMT-2785) have the majority of their shared IBDs with later Avar-period individuals (see also **Figure 19** and **IV.2.1**). This pattern provides further evidence that eastern immigrants distinct from the Steppe Sarmatians also appeared during these periods.

IV.5.2 IBD connections between period groups

Next, we analysed pairwise combinations of group connectedness across subsequent time frames (**Figure 13**). In this analysis the number of connections were taken into account instead of the total IBD length. We carefully normalized the degree centrality data by dividing the detected connections by the total number of possible connections, resulting in the ratio of fulfilled connections as described in **III.2.6**. The X-axis in **Figure 13** displays the short label of each group, while the columns represent the ratio of fulfilled IBD connections between the indicated group and every other group.

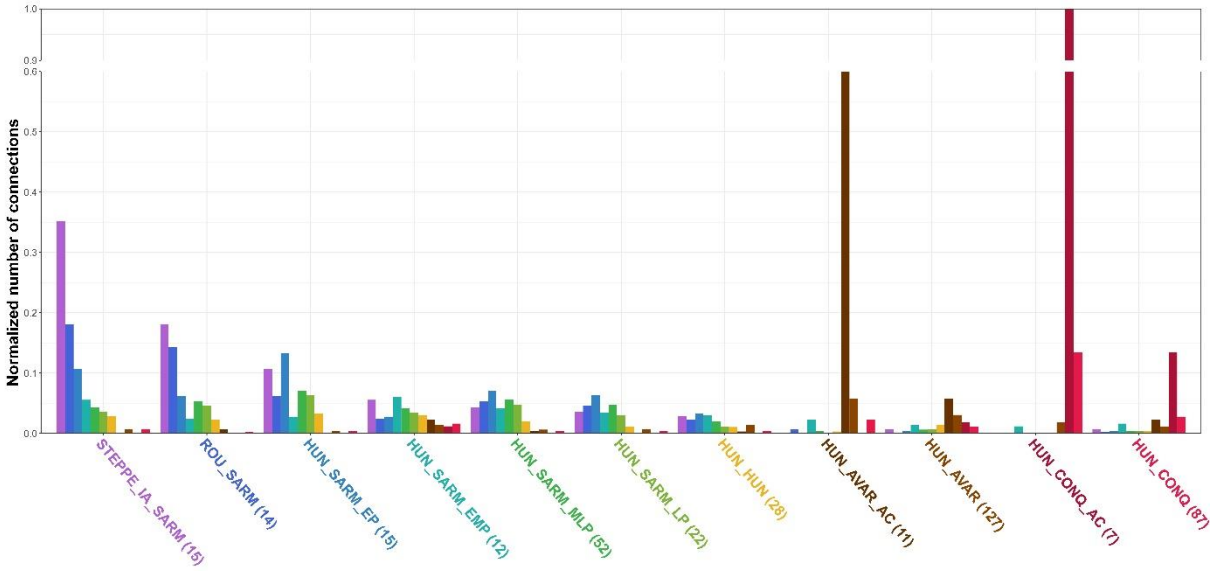


Figure 13: Normalized number of IBD connections across the different archaeological periods of the Carpathian Basin and its vicinity. The colours of the columns correspond to the colours of the group labels, with the height of each column representing the strength of the IBD connections for the group labelled with the letter code. The groups are arranged from left to right in chronological order. The plot has been truncated above the 0.6 line to accommodate the unusually high (100%) intragroup sharing of the HUN_CONQ_AC group. Normalized values were calculated as described in **III.2.6**.

Figure 13 illustrates that the Steppe Sarmatians (STEPPE_IA_SARM) exhibit a consistently decreasing sharing pattern across the progressive time periods of the Carpathian Basin. This trend is consistent with a possible founding effect, where the genomic contribution of the earliest group naturally diminishes over subsequent generations. Additionally, the steadily declining pattern of intergroup connections suggests a continuous chain of generational transmission, without any abrupt population turnovers throughout the successive archaeological periods.

A similar trend is observed in the connectedness of the ROU_SARM and HUN_SARM_EP groups with subsequent periods. However, there is a notable sharp decline in their connectedness with the HUN_SARM_EMP group, indicating a significant gap in IBD transmission during the Early-Middle Sarmatian Period. The HUN_SARM_EMP group again shows a declining pattern of IBD connections with subsequent periods but also reveals extensive connections with the post-Sarmatian Avar and Conquest-period groups (HUN_AVAR, HUN_CONQ), which were negligible in the earlier Sarmatians.

This phenomenon is likely attributed to a second wave of immigration during the EMP period, involving new groups. The HUN_SARM_EMP group is represented by two large cemeteries, Makó-Igási Járandó (MIJ) and Hódmezővásárhely-Fehértó (HVF). F4 and qpAdm analyses identified at least two individuals from MIJ (MIJ-1, MIJ-3) as potential migrants from Eastern or Central Asia. In contrast, four individuals from HVF (HVF-4, HVF-8, HVF-10, HVF-21) showed significant Northern European-related ancestry (see **Figure 10**), with 3 of these individuals being genetic outliers, and we were unable to model HVF-10 accurately (see **IV.4.2.3**). This suggests that the HVF population likely represents new migration from Northern Europe. Nevertheless, the MIJ and HVF cemeteries do not fully represent the entire population of the HUN_SARM_EMP period, as the populations from the HUN_SARM_MLP and LP periods show a much stronger connection with the HUN_SARM_EP group.

The HUN_HUN group exhibits the lowest level of IBD sharing within itself, instead showing stronger genetic connections to earlier periods. It also acts as a genetic bridge between the HUN_SARM_LP and HUN_AVAR groups, displaying an increased number of shared connections with the latter. It must be noted that many of the individuals classified into this group came from solitary graves or represent their cemetery alone. As a result, the observed intragroup sharing may underestimate the true degree of genetic connectedness within this period.

The so-called "immigrant cores" of the later Avar and Conquest Periods (HUN_AVAR_AC and HUN_CONQ_AC) described by Maróti et al. (2022), exhibit distinctive IBD sharing patterns compared to other groups, with the exception of the STEPPE_IA_SARM. Their prominent intragroup IBD sharing and relatively low sharing with contemporary neighbours indicate a distinct, endogamous population. In contrast, the majority of sequenced individuals from the Avar and Conquest Periods (HUN_AVAR, HUN_CONQ) exhibit a more regular sharing pattern, suggesting they likely represent a broader segment of the population from that era. Despite a reduction in connections, links to Sarmatian-period individuals are still observable in these later groups, indicating that at least a portion of the pre-Hun-period population persisted in the Carpathian Basin.

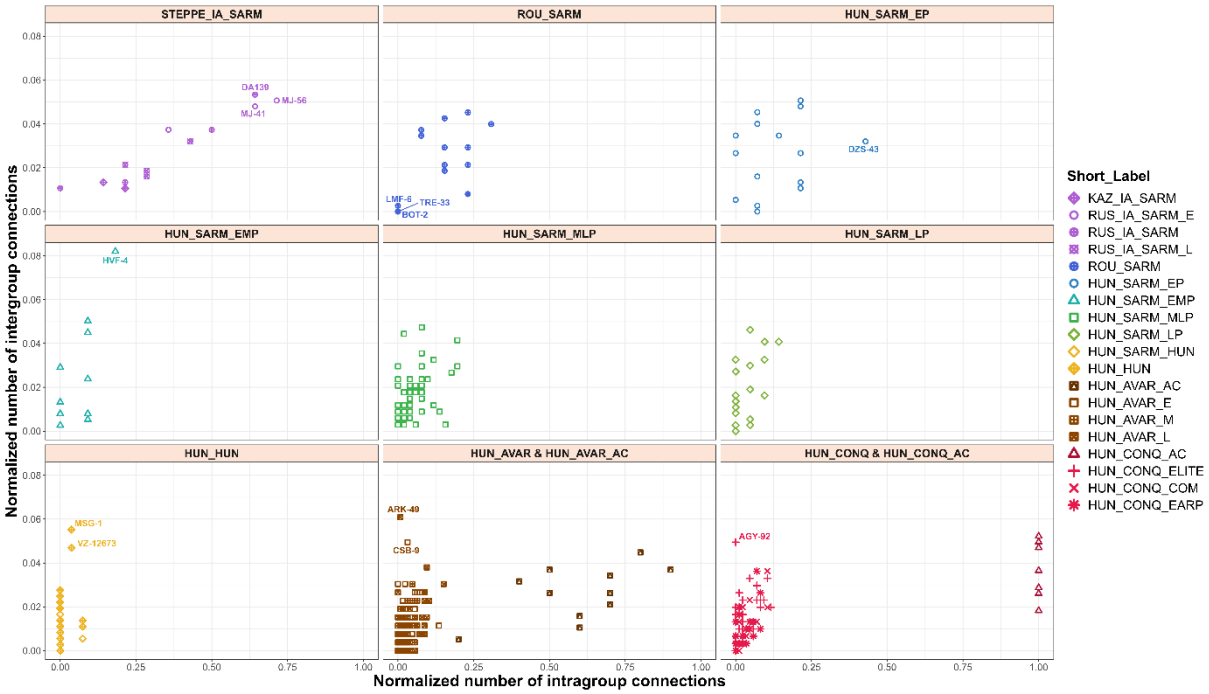


Figure 14: Normalized counts of intragroup and intergroup connections of individuals across various archaeological periods in the Carpathian Basin and its vicinity. Group names are written in the strip text above each subplot. For simplicity, the HUN_AVAR-HUN_AVAR_AC and HUN_CONQ-HUN_CONQ_AC groups are plotted together, although calculations were performed using the original groupings. The Y-axis represents the normalized number of intergroup connections (number of detected outgroup connections/total number of all individuals - group size]). The X-axis represents the normalized number of intragroup connections (number of detected intragroup connections/[group size – 1]). Each connection was considered equal, independent of IBD number or size.

These results are supplemented by the individual connectivity plotted across each group (**Figure 14**). The X axis represents the normalized number of intragroup connections, while the Y axis shows the normalized number of intergroup connections. A very interesting pattern is apparent as 3 groups (STEPPE_IA_SARM, HUN_AVAR_AC and HUN_CONQ_AC) display unusually high ratios of intragroup sharing compared to the others. This pattern can be also observed in the ROU_SARM and HUN_SARM_EP groups to a certain extent but declines steadily over the progressive periods. A possible explanation for this phenomenon could be the effect of nomadism which facilitate the formation of family bounds through elevated mobility across large geographical distances, while the agricultural lifestyle tends to localise people, leading to sparser genetic connectedness even across relatively short distances. Based on archaeological and historical research, the Carpathian Basin Sarmatians rapidly adapted to their new environmental conditions. While livestock farming still remained one of their key economic drives, they gradually incorporated agriculture and subsistence farming (see **I.1.3**). This transition may be reflected in the gradual decline of intragroup connectedness observed in later Sarmatian groups. In contrast, the Steppe Sarmatians and later steppe migrants continued to exhibit high levels of genetic relatedness within their groups. While factors such as population growth or sampling bias could also contribute to this trend, the elevated intragroup connectedness should be considered a potential marker of nomadism, at least in the case of the STEPPE_IA_SARM group. These individuals come from a very wide spatio-temporal range and still display this unusual pattern.

An intriguing case is HVF-4, a Northern European genetic outlier identified by its PCA position and qpAdm models (labelled within the HUN_SARM_EMP group on **Figure 14**, see also **Figures 5** and **10** and **IV.4.2.3**). This individual exhibits the highest number of genealogical connections within the entire dataset, including ancient individuals not depicted in **Figure 14**, harbouring links to 36 different individuals (~1.74% of all possible connections). Notably, HVF-4 is connected to multiple Avar and Conquest-period individuals, including some from the Conquest-period elite. While this could simply be a case of a lucky individual with an unusually high number of surviving descendants, it is remarkable that we were able to identify someone with such demographic significance. This finding also reinforces the possible persistence of Sarmatian-period lineages up until the Hungarian Conquest Period.

Individuals from the HUN_HUN group exhibit the lowest levels of intragroup connectivity, with the vast majority showing no links to other members of their group. However, their intergroup connectivity is quite high, particularly in the cases of MSG-1 and VZ-12673, two

East Asian immigrants who display an above-average number of connections within the context of our dataset. The combination of low internal connectivity and high external connectivity further highlights the artificial nature of this population grouping.

IV.5.3 IBD connections between the sampled cemeteries

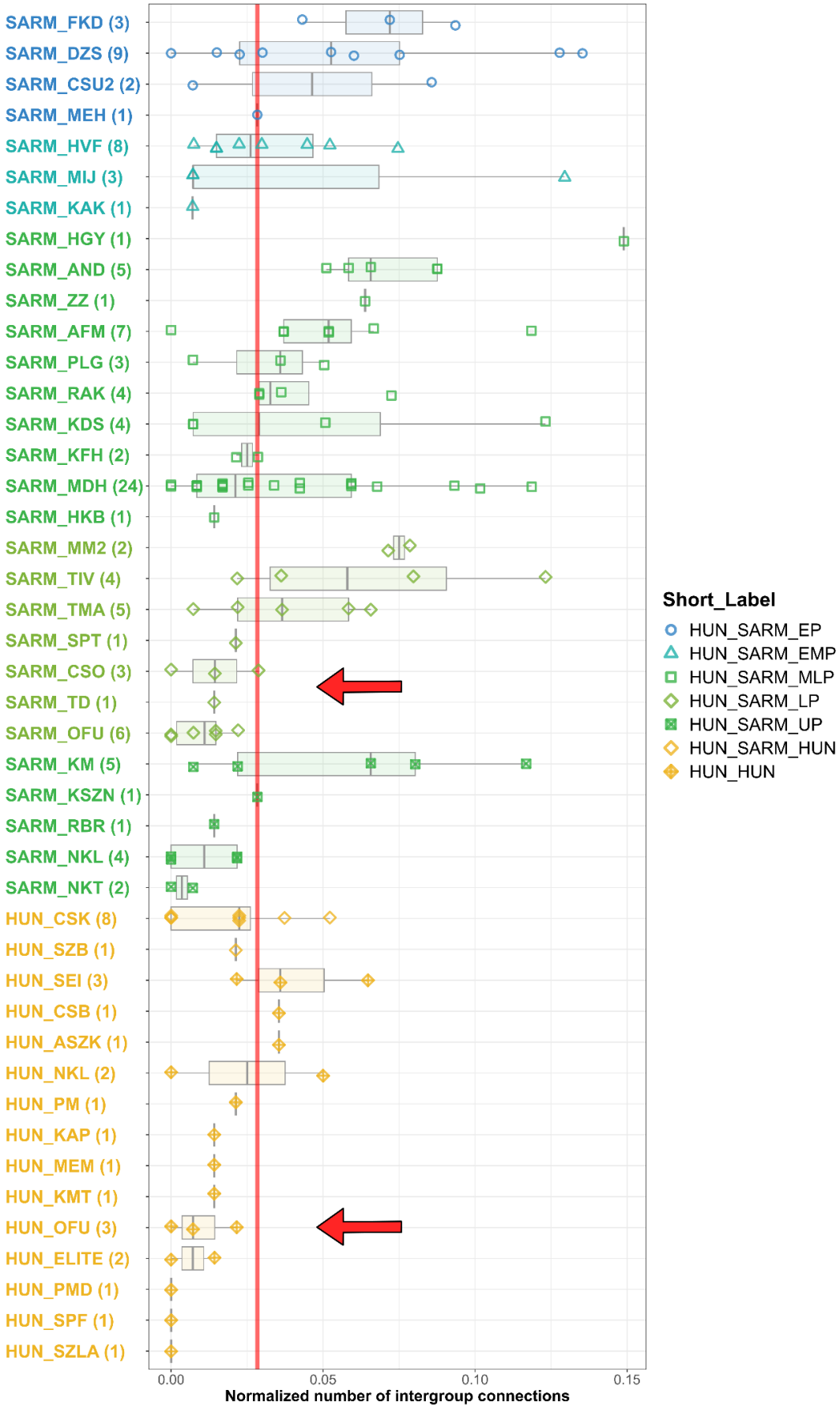


Figure 15: Intergroup sharing among cemeteries of the Sarmatian and Hun Periods of the Carpathian Basin. The red line indicates the median of the intergroup IBD connections across the plotted individuals. Cemetery labels were created (see **Table E3a**) by combining the sample's archaeological period with the three-letter cemetery code from the sample's Master ID (for further details see **III.2.6**).

Plotting intergroup IBD sharing by cemetery (**Figure 15**) reveals a significant depletion of IBD connections in about half of the Late Sarmatian (HUN_SARM_LP) and Hun-period (HUN_HUN) cemeteries (red arrows), compared to earlier periods. Several possible explanations could account for these results. The first is sampling bias, as many of the individuals with low connections are from solitary finds or represent their cemetery alone, suggesting that more connections might emerge with more comprehensive sampling. The second explanation could be an increasing population size, combined with sampling bias. A third possibility is that these cemeteries primarily represent new migrants with limited connections to the preceding local population.

In the case of the Hun Period, we have documented migrants with very strong Inner Asian genetic connections, suggesting that migration is the most plausible explanation, even if sampling bias is a factor. Besides, archaeologists detect considerable population decline in the Hun Period, so an increasing population size is an unlikely factor.

Migration is also the most likely explanation for the same phenomenon in the Late Sarmatian Period, especially when considering the genomic composition of individuals with low IBD connections. These individuals display sparse IBD ties to the Sarmatians, but show significant connections to the later Avar Period, which was not considered in **Figure 15**. This is particularly evident in the Óföldeák–Ürmös (OFU) cemetery from the HUN_SARM_LP group, where three individuals (OFU-168, OFU-190, OFU-422) have close kinship relations with both published and unpublished Avar-period individuals (see **Table S4**). Interestingly, these Avars were not Asian immigrants but originated from Europe. Similarly, most individuals from the HUN_SARM_LP group with low connections (indicated by the red arrow on **Figure 15**) show a clear PCA shift toward modern Southern Europeans. In qpAdm analysis, these individuals exhibit the strongest affinities with sources related to the Roman Empire (e.g., Austria_Ovilava_Roman.SG, Italy_Imperial.SG, Germany_Roman.SG, Italy_IA_Republic.SG), along with some local and Sarmatian admixture (**Table E2** and **Figure 10**). Therefore, in the Late Sarmatian Period, the new migration likely originated from neighbouring Roman provinces rather than the steppes.

IV.5.4 IBD sharing network of selected cemeteries and period groups

To better understand the fine scale demographic processes of the Carpathian Basin Sarmatians throughout the centuries of their occupation, we generated separate IBD graphs highlighting the genealogical links between key cemeteries and groups. These graphs were constructed by selecting individuals from specific cemeteries or groups and supplementing them with all other individuals who shared a minimum total IBD length of 12 cM with at least one of the selected samples (with a minimum individual IBD segment length of 8 cM). However, edges corresponding to <12 cM IBD lengths were still retained if they appeared between the selected individuals. This was primarily done to generate comprehensible plots as large cemeteries would become overcrowded otherwise. Furthermore, this approach emphasizes closer genealogical connections and thus provide an insight into possible personal-level genealogies.

IV.5.4.1 The central position of the HUN_SARM_EP group

Figure 16 shows the IBD connection network of the HUN_SARM_EP group. As we saw on **Figure 12**, they exhibit possible signs of a founder effect. This is further supported by their extensive connections, both with Steppe Sarmatians and individuals from later periods. Individuals from Dormánd – Zsidótemető seem to represent a close-knit group as almost all sampled individual belongs to an extended family.

In contrast, the three individuals from Füzesabony – Kastélydűlő show much less connections among each other, rather they present three distinct profiles. FKD-60 shares numerous short IBD connections with individuals from the HUN_SARM_EMP and HUN_SARM_MLP groups, as well as with a HUN_SARM_HUN individual from a period potentially three centuries later. FKD-150 - a genetically female individual - exhibits a different pattern, sharing IBD connections almost exclusively with Steppe Sarmatians from the Southern Ural region and members of the ROU_SARM group. Especially interesting is her connection to DA139 (another Steppe Sarmatian female from the Caspian Steppe), with whom she shares a total lengths of 48 cM IBDs in 4 fragments, indicating a closer than 10-degree genealogical connection. FKD-140 occupies an intermediate position as it has shared IBDs both with later HUN_SARM_MLP and Steppe Sarmatian individuals.

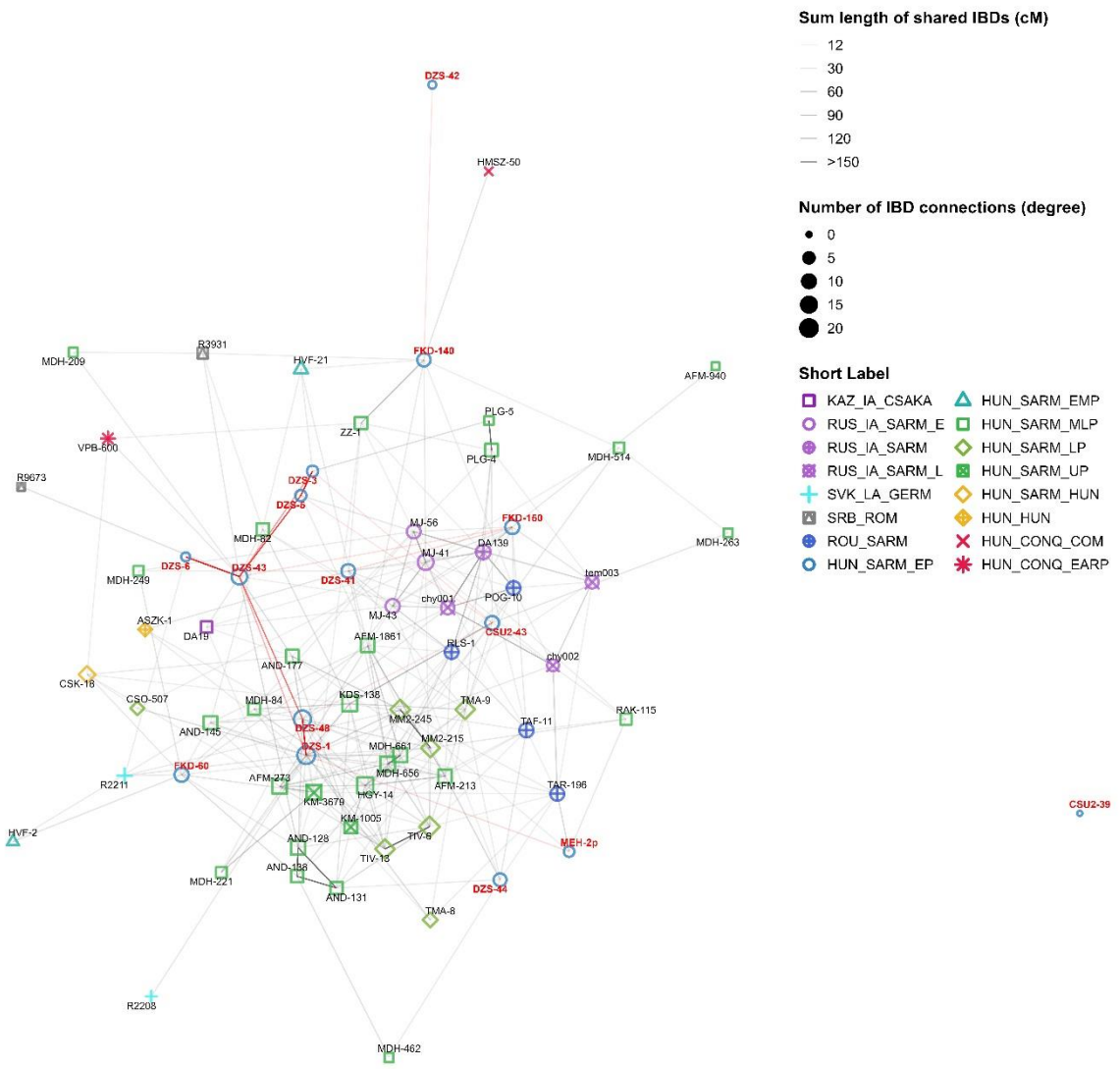


Figure 16: IBD graph of the HUN_SARM_EP individuals indicated with red labels (prepared according to **IV.5.4)**. Intragroup connections are signified by red edges. For details on the Short Label abbreviations see **Table S5**.

These findings, in conjunction with the qpAdm analysis results (see **IV.4.2.1**), strongly support the hypothesis that the HUN_SARM_EP individuals directly descended from Steppe Sarmatians who migrated into the Carpathian Basin. Moreover, their extensive connections to individuals from subsequent periods further reinforce their role as a likely source of the Steppe Sarmatian ancestry detected in many of the studied individuals.

IV.5.4.2 Demographic landscape in the Middle-Sarmatian Period

The largest sampled cemetery in our study is from Madaras – Halmok, primarily representing the population of the Middle-Late Sarmatian Period (**Figure 17**). A notable difference compared to the HUN_SARM_EP individuals is the sparse intra-cemetery connectivity. Out of a possible 276 connections, only 9 were detected within this group. This low degree of internal relatedness may suggest that the cemetery belonged to a more metropolitan population centre (also indicated by its large size). However, sampling bias could also be a contributing factor, as we were only able to analyse a fraction of the excavated individuals due to their poor preservation quality.

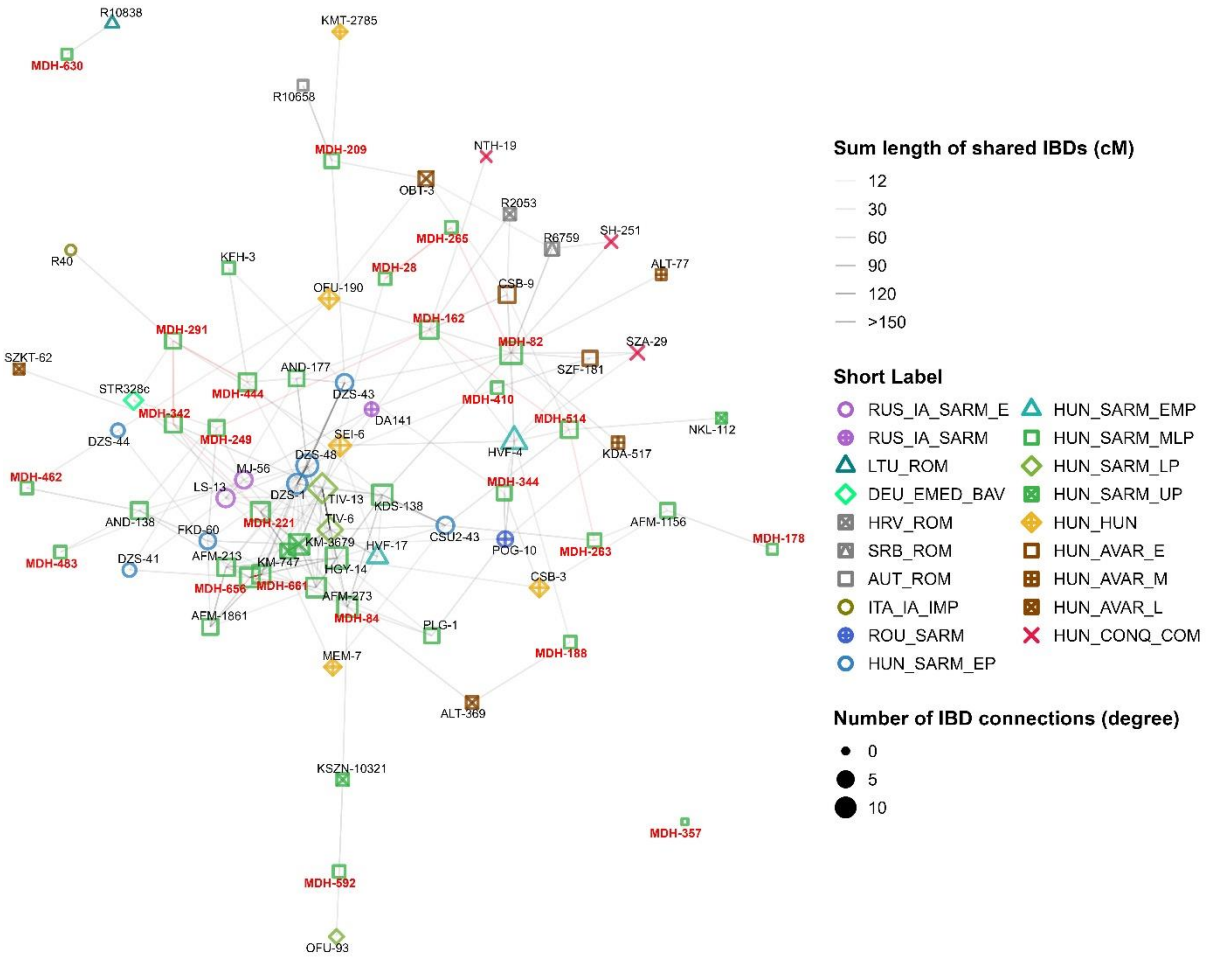


Figure 17 IBD graph of the individuals from Madaras – Halmok (prepared according to **IV.5.4)**. SARM_MDH individuals are all plotted with red labels indicating their identity, intragroup connections are signified by red edges. For details on the Short Label abbreviations see **Table S5**.

Despite the limited intra-cemetery connections, most of the outgoing IBD links place the individuals among the other Sarmatians from the region, highlighting the broader interconnectedness of the population. Several notable genealogies were identified, including strong connections to Dormánd – Zsidótemető (14 connections) from the Early Period, as well as to Apc – Farkas-major (9 connections) and Tiszavalk (8 connections) from the Middle and Late Sarmatian Periods. Interestingly, these sites are located hundreds of kilometers to the north, near the foothills of the Northern Mountains of Hungary. These results may indicate considerable mobility within the community living on the Great Hungarian Plains, although their overall connectedness declined as seen on **Figure 14**.

IV.5.4.3 Population changes during the Migration Period

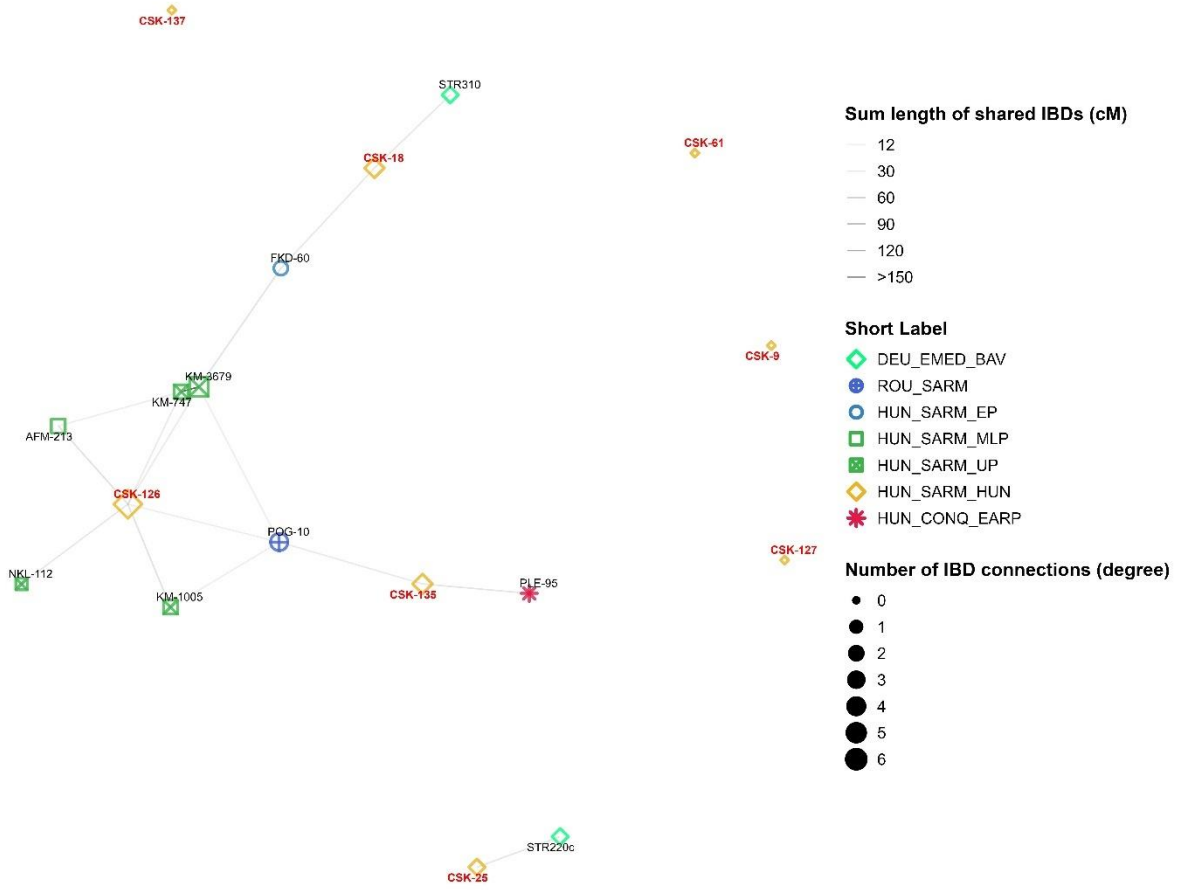


Figure 18: IBD graph of the individuals from Csongrád – Kenderföldek (prepared according to **IV.5.4)**. Red labels indicate the HUN_CSK individuals, intragroup connections were not detected in this group. For details on the Short Label abbreviations see **Table S5**.

We can look at the supposed transition between the Late Sarmatian and Hun Period at the turn of the 4th century through the cemetery Csongrád – Kenderföldek (**Figure 18**). This cemetery

is quite unique in our dataset as it appears to have been in continuous use from the Late Sarmatian Period through the Hun Period. Interestingly, we could not identify any IBD connections among the individuals found in this cemetery.

A closer examination reveals that the individuals from Csongrád – Kenderföldek can be categorized into two distinct groups. Some individuals (e.g., CSK-18, CSK-126, and CSK-135) exhibit clear genetic connections to earlier periods, while others (like CSK-61 or CSK-137) completely lack any embedded relations in the region. This may be an indicator of a mixed community living in this location, with some individuals with entrenched history in the region and others arriving recently into the Carpathian Basin.

Notably, we also identified links between individuals from this cemetery and Early Medieval Bavaria, specifically with individuals from the Straubing – Bajuwarenstraße cemetery (e.g., STR220c and STR310). This connection is particularly interesting, as the Straubing site has previously been associated with possible Sarmatian migrants that arrived during the Migration Period (Geisler, 1998; Martin, 1995).

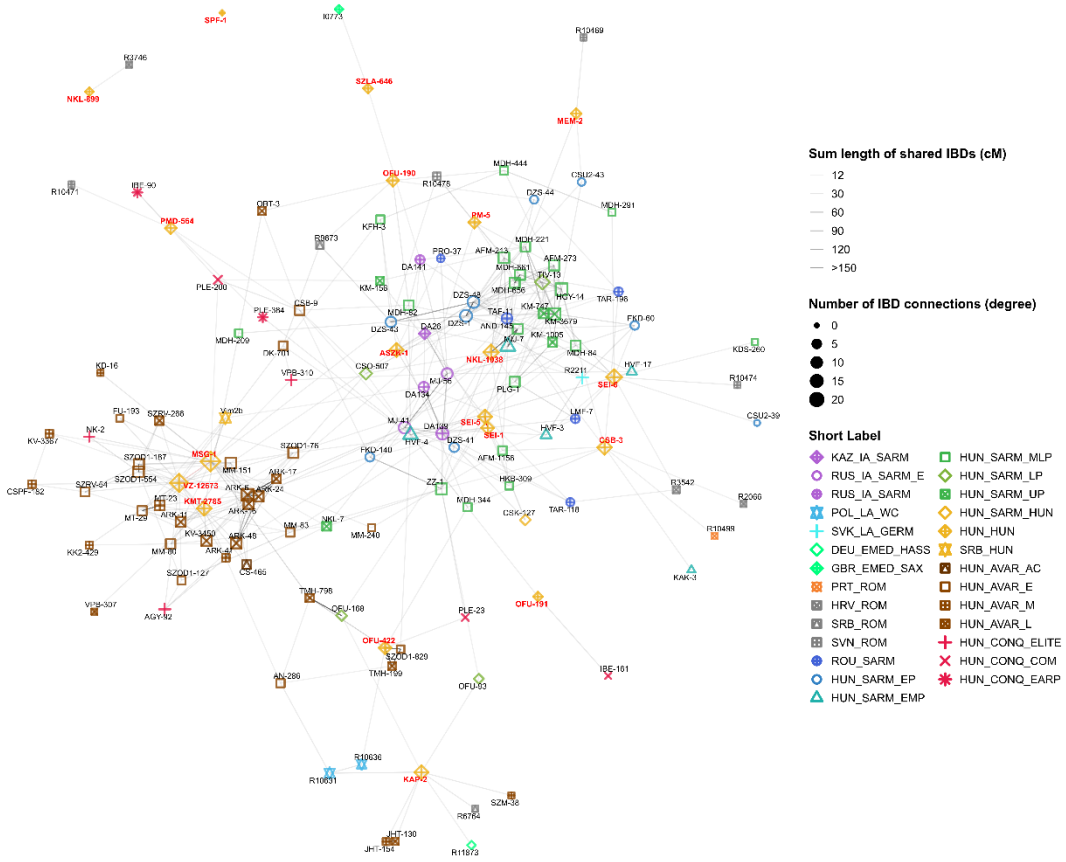


Figure 19: IBD sharing graph of the HUN_HUN individuals indicated with red labels (prepared according to **IV.5.4)**. Intragroup connections are highlighted with red edges. For details on the Short Label abbreviations see **Table S5**.

Analysing the IBD sharing patterns of Hun-period individuals offers valuable insights into the demographic shifts that followed the Great Migration Period. As seen in **Figure 19**, a pattern similar to that observed in the Csongrád – Kenderföldek cemetery emerges.

The HUN_HUN samples are distinctly separated into two IBD-sharing clusters. One cluster primarily consists of Avar-period samples, including a few Conqueror elites. The other cluster is dominated by Sarmatian-period individuals, including several Romanian and Steppe Sarmatians, with only marginal connections to Roman, Avar, and Conquest-period individuals. The Hun-period samples within the Avar cluster are predominantly characterized by Asian genomic components (see **Figure 10** and **IV.4.2.2**), indicating they represent recent immigrants from Asia. In contrast, the Hun-period samples within the Sarmatian cluster clearly descend from the local Sarmatian-period population.

This suggests that while significant cultural shifts occurred in the region - evidenced by the emergence of new archaeological artifact types - the descendants of the earlier Sarmatian populations continued to inhabit the Great Hungarian Plain. However, the presence of individuals with no apparent genealogical background in the region, along with those exhibiting distinct genetic compositions, may indicate the vectors of the observed cultural transition while also pointing to a broader, continent-wide migration event.

IV.5.5 IBD analysis of the Steppe Sarmatians

IV.5.5.1 IBD connection network of the Steppe Sarmatians

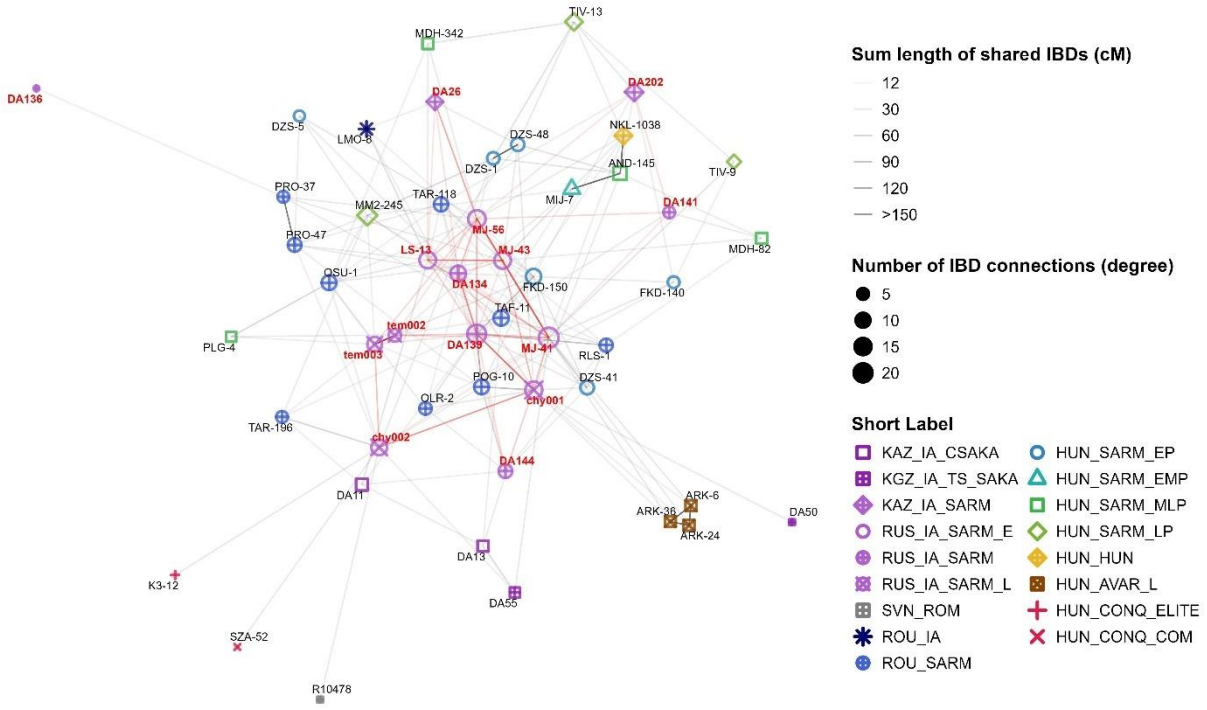


Figure 20: IBD connection network of the Steppe Sarmatians highlighted with red labels (prepared according to **IV.5.4)**. Connections among Steppe Sarmatians are indicated with red edges. For details on the Short Label abbreviations see **Table S5**.

Figure 20 illustrates the IBD sharing network of the Steppe Sarmatian individuals available in the literature. They share most of their connections with the newly sequenced Sarmatians especially with those from Romania and the Early Sarmatian Period of the Carpathian Basin. Additionally, several connections are observed with individuals from the Carpathian Basin Avar and Conquest Periods, as well as a Roman individual from present-day Slovenia. Notably, LMO-8, an Iron Age individual from Romania, also appears in the network with multiple connections.

As seen in **Figure 11**, the Steppe Sarmatians exhibit minimal connections to publicly available Western Scythians, despite their shared history attested by the ancient authors (Istvánovits & Kulcsár, 2017). Instead, they display multiple connections with Iron Age individuals associated with the Central Asian Saka culture (marked with purple in **Figure 20**). This reinforces the findings of Gnecchi-Ruscone et al. (2021), which highlighted strong genetic links between Sarmatians from the Ural region and Central Steppe nomads (see also **I.2.3**).

Table 3 further demonstrates that Eastern Scythians harbour numerous genetic connections with Sarmatians, not only with Steppe Sarmatians from the Ural region but also with the newly sequenced Sarmatians from Romania and the Carpathian Basin. These connections, spanning multiple centuries, provide further evidence of a direct genetic link between Carpathian Basin Sarmatians and those from the Ural region.

Table 3: IBDs shared between the Eastern Scythians and Sarmatians available in our database. Colours were given according to the colour code previously used in all figures (purple – Eastern Scythians, lavender – Sarmatians, blue – Iron Age and Sarmatian samples from Romania, green – Sarmatians from the Carpathian Basin). For details on the Short Label abbreviations see **Table S5**.

Master ID 1	Master ID 2	Short Label 1	Short Label 2	Sum of IBDs (cM)	Number of IBD segments
DA11	LS-13	KAZ_IA_CSAKA	RUS_IA_SARM_E	19.19	2
DA19	DZS-43	KAZ_IA_CSAKA	HUN_SARM_EP	14.53	1
DA55	MJ-41	KGZ_IA_TS_SAKA	RUS_IA_SARM_E	13.95	1
DA13	MJ-41	KAZ_IA_CSAKA	RUS_IA_SARM_E	13.39	1
DA50	chy001	KGZ_IA_TS_SAKA	RUS_IA_SARM_L	12.95	1
DA53	LMO-8	KGZ_IA_TS_SAKA	ROU_IA	12.33	1
DA10	POG-10	KAZ_IA_CSAKA	ROU_SARM	11.52	1
DA56	MJ-41	KGZ_IA_TS_SAKA	RUS_IA_SARM_E	10.70	1
DA49	MJ-41	KGZ_IA_TS_SAKA	RUS_IA_SARM_E	10.41	1
DA55	chy002	KGZ_IA_TS_SAKA	RUS_IA_SARM_L	10.31	1
DA13	TAF-11	KAZ_IA_CSAKA	ROU_SARM	9.86	1
DA19	MJ-43	KAZ_IA_CSAKA	RUS_IA_SARM_E	9.59	1
DA19	AND-128	KAZ_IA_CSAKA	HUN_SARM_MLP	9.25	1
DA49	MDH-444	KGZ_IA_TS_SAKA	HUN_SARM_MLP	9.01	1
DA17	LMO-8	KAZ_IA_CSAKA	ROU_IA	8.98	1
DA11	DA144	KAZ_IA_CSAKA	RUS_IA_SARM	8.91	1
DA13	FKD-60	KAZ_IA_CSAKA	HUN_SARM_EP	8.90	1
DA11	TAR-196	KAZ_IA_CSAKA	ROU_SARM	8.74	1
DA53	DA26	KGZ_IA_TS_SAKA	KAZ_IA_SARM	8.69	1
DA19	AND-177	KAZ_IA_CSAKA	HUN_SARM_MLP	8.69	1
DA19	MJ-41	KAZ_IA_CSAKA	RUS_IA_SARM_E	8.69	1
DA53	KM-747	KGZ_IA_TS_SAKA	HUN_SARM_UP	8.46	1
DA53	KM-3679	KGZ_IA_TS_SAKA	HUN_SARM_UP	8.45	1
DA11	DA139	KAZ_IA_CSAKA	RUS_IA_SARM	8.40	1
MJ-42	TAF-11	RUS_IA_SCY	ROU_SARM	8.12	1

Table 4 summarizes all shared connections between the available Western Scythians and Sarmatians across different periods. A notable difference can be seen between **Tables 3** and **4**, as the Sarmatians exhibit significantly more connections with the Central Steppe Scythians than with the Western Scythians of the Caspian Steppe and the Carpathian Basin.

This finding is particularly intriguing given that historical sources, including Herodotus, describe the Sarmatians as vassals of the Royal Scythians of the Pontic region (Istvánovits & Kulcsár 2017). Despite this historical relationship, genetic links between the two groups appear to be minimal. Even more surprising is the lack of strong genetic ties between the Western Scythians and the Carpathian Basin Sarmatians, despite inhabiting the same geographical region.

Furthermore, the majority of the detected connections are concentrated around a single individual, MJ-16, further underscoring the apparent genetic isolation of this group. While the age difference and potential sampling bias may contribute to this pattern, it is notable that the Central Steppe Scythians are not overrepresented in the dataset, on the contrary their group comprises only 12 individuals whereas the Western Scythians number 17. This suggests that the observed pattern is not merely a product of uneven sampling but may reflect a genuine demographic phenomenon.

Table 4: IBDs shared between the Western Scythians and Sarmatians available in our database. Colours were given according to the colour code previously used in all figures (light lavender – Western Scythians, dark lavender – Sarmatians, blue – Iron Age and Sarmatian samples from Romania, green – Sarmatians from the Carpathian Basin). For details on the Short Label abbreviations see **Table S5**.

Master ID 1	Master ID 2	Short Label 1	Short Label 2	Sum of IBDs (cM)	Number of IBD segments
MJ-16	OSU-1	UKR_IA_SCY	ROU_SARM	11.66	1
MJ-16	LMF-7	UKR_IA_SCY	ROU_SARM	11.30	1
MJ-16	LS-13	UKR_IA_SCY	RUS_IA_SARM_E	10.02	1
MJ-16	LMO-8	UKR_IA_SCY	ROU_IA	9.56	1
scy009	HVF-10	UKR_IA_SCY	HUN_SARM_EMP	9.09	1
MJ-16	MJ-56	UKR_IA_SCY	RUS_IA_SARM_E	8.81	1
MJ-16	AFM-273	UKR_IA_SCY	HUN_SARM_MLP	8.65	1
MJ-16	chy001	UKR_IA_SCY	RUS_IA_SARM_L	8.58	1
scy301	RAM-7	MDA_IA_SCY	ROU_IA	8.28	1
scy009	MDH-84	UKR_IA_SCY	HUN_SARM_MLP	8.18	1

The presence of LMO-8 in the IBD network (**Figure 20**) is particularly intriguing. As the oldest sample in our dataset (dated to 801–753 calBCE), LMO-8 exhibits a genetic profile closely resembling that of the Steppe Sarmatians, albeit with a slightly higher proportion of East Asian ancestry (see **IV.2.2** and **IV.4.1**). Their IBD connections offer further valuable insights (**Table 5**).

LMO-8 shares links with multiple Steppe Sarmatians, as well as with Saka individuals from the Tian Shan and Central Steppe regions, a Western Scythian from Ukraine, and a member of the HUN_SARM_EP group. Most striking is their connection with LS-13, sharing two IBD fragments spanning two chromosomes (one on chr-1 and another on chr-13), suggesting a direct genealogical relationship despite the ~300-year time gap between them. This finding further reinforces the direct association of this early individual with the Steppe Sarmatians while also highlighting a tangible link between the Scythian cultures of the Western and Eastern Steppes.

Table 5: IBD connections of LMO-8, an Iron Age individual excavated in Romania, the oldest sample in our dataset. Colours were given according to the colour code previously used in all figures (light lavender – Western Scythians, dark lavender – Sarmatians, purple – Eastern Scythians, green – Sarmatians from the Carpathian Basin). For details on the Short Label abbreviations see **Table S5**.

Master ID	Short Label	Sum of IBDs (cM)	Number of IBD segments
LS-13	RUS_IA_SARM_E	22.50	2
DA53	KGZ_IA_TS_SAKA	12.33	1
MJ-16	UKR_IA_SCY	9.56	1
DA17	KAZ_IA_CSAKA	8.98	1
DA139	RUS_IA_SARM	8.68	1
FKD-150	HUN_SARM_EP	8.35	1
DA26	KAZ_IA_SARM	8.04	1

IV.5.5.2 An extended Sarmatian family network across the Eurasian Steppe

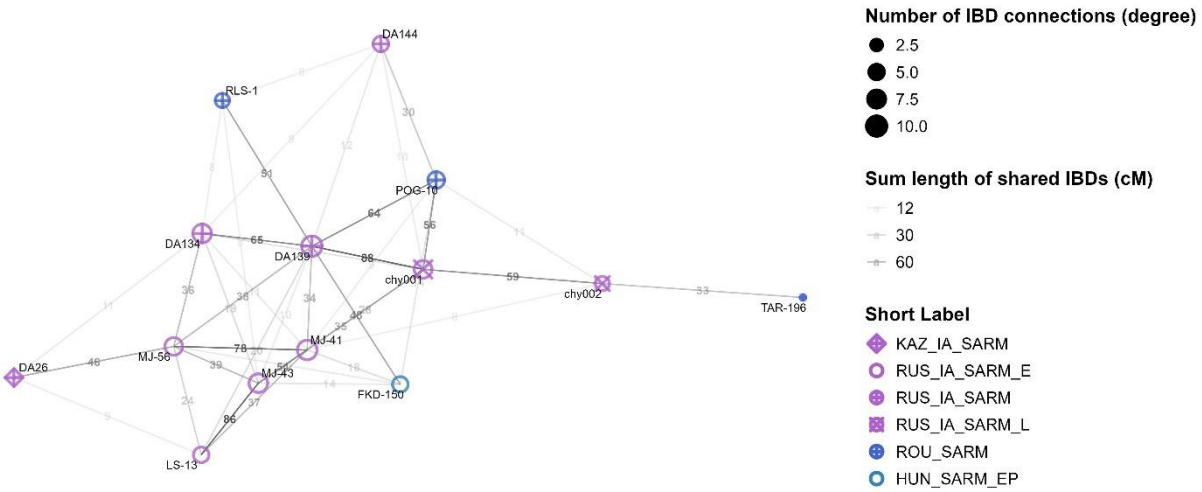


Figure 21: Extended Steppe Sarmatian family including a ROU_SARM and a HUN_SARM_EP individual. The family was identified by selecting individuals who share connections with a minimum total IBD length of 30 cM across at least two segments.

One of the most remarkable findings is the presence of an extended Sarmatian family network (**Figure 21**) spanning almost the entire Eurasian Steppe, from the Central Steppe region (DA26) to the northern reaches of the Great Hungarian Plain (FKD-150). This transcontinental genealogical connection network, assembled from multiple publications (4 articles including our own), defies expectations and highlights the deep-rooted kinship ties among the Sarmatian populations.

The term "extended family" is used rather loosely in this context, as it is unlikely that these individuals knew each other personally. However, these connections should not be interpreted as the result of isolated, short IBD segments that can persist over many centuries. Rather, the observed IBD connections reflect close relations that still can be understood as genuine familial ties (<10th degree of relatedness, probably on a horizontal scale).

This relative group includes several individuals from the Early Sarmatian Period of the Steppe, as well as three Romanian Sarmatians (POG-10, RLS-1, and TAR-196) and an individual from the Early Sarmatian Period of the Carpathian Basin (FKD-150). Their shared ancestry confirms a direct genealogical link between the Sarmatians of the Carpathian Basin and those of the Southern Ural region.

The network consists of two primary family subunits. One includes Early Sarmatians from the Southern Ural region (MJ-41, MJ-43, MJ-56 and LS-13), published in Järve et al. (2019), while

the other comprises three individuals excavated near the Azov Sea (DA134, DA139, DA144 published in Damgaard et al., 2018), two Sarmatians from the Ural region (chy001 and chy002 published in Krzewińska et al., 2018), and the Romanian and Carpathian Basin Sarmatians. The first group appears to be balanced and contains LS-13, a key individual exhibiting multi-segment IBD connections with individuals across a vast spatio-temporal range (see **Tables 3, 4,** and **5**). The second group is heavily centred around DA139, a female individual excavated at the Don Delta. Interestingly, this extended family exhibits a disproportionate number of females, with only 5 males (DA144, LS-13, chy002, POG-10, and TAR-196) compared to 9 females. Furthermore, the males appear to occupy peripheral positions within the network, whereas the females show a significantly higher degree of interconnectedness.

The primary factor distinguishing these subgroups is likely their chronology. The four Early Sarmatians from Järve et al. (2019) span a potentially broad temporal range but are generally dated to around 400 BCE. The second group most likely represents a later contemporaneous cluster. Two of the Sarmatians from Romania (POG-10 and TAR-196) are dated to the 1st century BCE while the two other Sarmatians from the Ural region (chy001 and chy002), along with the Early Period Sarmatian from the Carpathian Basin (FKD-150), are dated to the 2nd century CE. The Azov Sarmatians lack precise dating, but their familial connections suggest they likely fall within this timeframe, around the 1st century CE. This would make them contemporaneous with the remaining Romanian Sarmatian (RLS-1), who is also dated to this period. This network, spanning both the temporal and spatial breadth of the Sarmatian era, provides compelling evidence for long-standing familial continuity across the Steppe.

IV.6 Uniparental data

Uniparental data, a specific form of IBD information, can provide valuable insights into population movements and familial relations. To leverage this, we assembled a comprehensive database of ancient individuals from the Carpathian Basin. The database was created with the AADR (Mallick et al., 2024) as its foundation and included samples from countries within the Carpathian range, as well as Steppe Sarmatian samples for reference. These countries include Austria, Slovenia, Croatia, Serbia, Romania, Slovakia, Hungary, and Russia and Kazakhstan for the Steppe Sarmatians.

The uniparental data was collected from original publications, although some samples were excluded due to insufficient data. For the final haplogroup classifications, we consulted databases published in Freeman et al. (2020) and Maár et al. (2021), as well as our own classifications for genomes already downloaded for IBD analysis (Antonio et al., 2024; Damgaard et al., 2018; Veeramah et al., 2018). We further excluded first-degree relatives, samples with a contamination rate higher than 10%, and low-coverage samples where mitochondrial haplogroups (Mt-Hg) could not be determined.

The final table includes 1,148 individuals, encompassing newly reported samples, spanning from the Mesolithic Period of the Carpathian Basin to the arrival of the Conquering Hungarians (**Table E4**). These individuals are divided into 512 females and 636 males.

IV.6.1 Repeated turnovers in the paternal haplogroup distribution

When examining the distribution of Y-chromosome haplogroups (Y-Hg) across different periods of the Carpathian Basin (**Figure 22**), we observe significant turnover in male lineages across multiple time periods. The Neolithic turnover linked to the migration of Anatolian farmers (Lipson et al., 2017; Mathieson et al., 2015, 2018) and the Bronze Age turnover associated with the Yamnaya migrations (Allentoft et al., 2015; Lipson et al., 2017; Mathieson et al., 2018) are well-documented. However, during the Sarmatian Period, we observe a new and previously unreported shift: the R1a~ haplogroups R-Z283 and R-Z93 appear in abundance and remain prevalent into the Hun Period. Particularly notable is the subclade R1a1a1b2~ (R-Z93), which is characteristic of Middle-Late Bronze Age Steppe populations such as the Sintashta and Andronovo cultures (Narasimhan et al., 2019), and their Iron Age descendants, the Eastern Scythians (Damgaard et al., 2018; Gnecchi-Ruscone et al., 2021; Wang et al., 2021). This haplogroup is also the most prevalent among the Steppe Sarmatians and Romanian

Sarmatians, highlighting a direct link between all Sarmatian groups and reinforcing the conclusions drawn from the autosomal data.

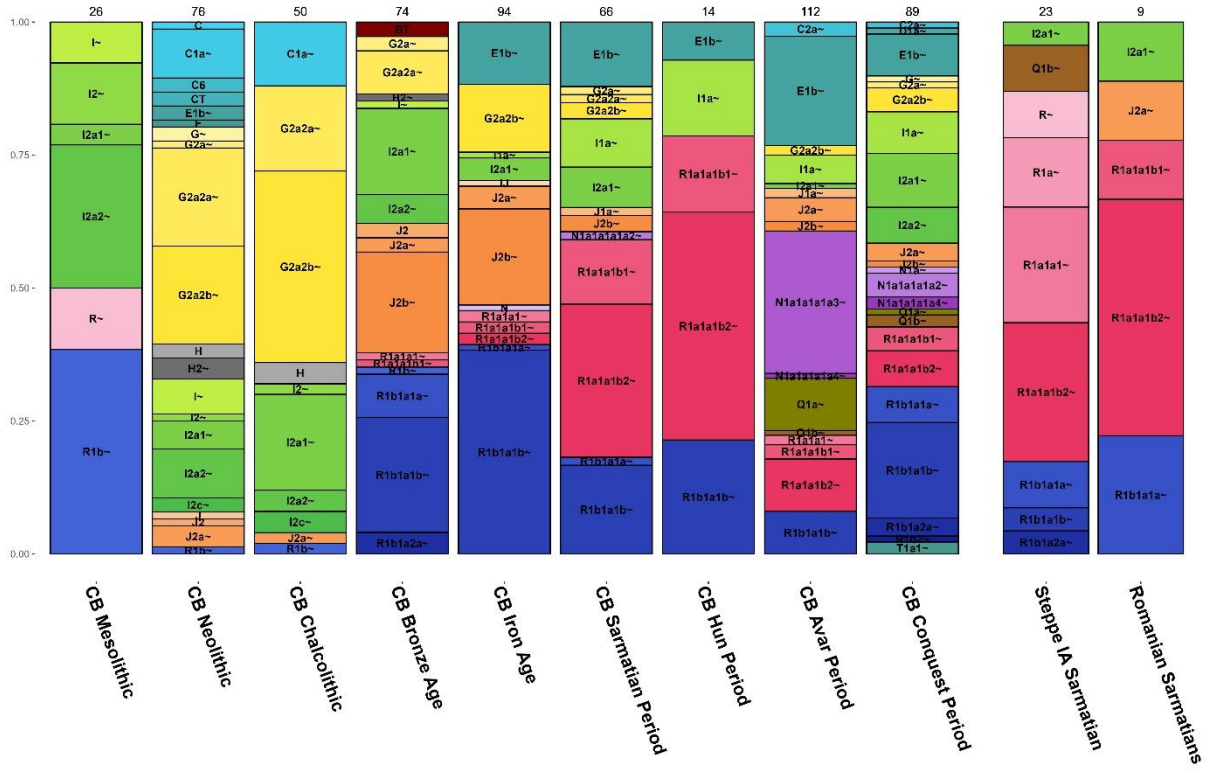


Figure 22: Y-haplogroup distribution of the progressive Carpathian Basin populations. The list of individuals contained in each period can be found in **Table E4**.

While the subclades of R1a~ were present only marginally in the Carpathian Basin before the appearance of the Sarmatians (1 Hungary_IA_Scythian [DA197], 1 Hungary_IA_LaTene_o3 [I25524] and 4 individuals from the Late Antiquity who were contemporaneous with the Sarmatians [R2211, POP23, R6759, R9673]), their prevalence is much higher among Western Scythians form Ukraine. This suggests that this Hg was already present in the vicinity of the Carpathian Basin during the Iron Age. However, several Ukraine_IA_WesternScythian individuals (MJ-14, MJ-15, MJ-33, MJ-34), who share similar R1a subhaplogroups with the Steppe Sarmatians, have been classified as genetic outliers. These individuals also exhibit very late radiocarbon dating (later than ~300 calBCE), which may indicate their direct connections to the Sarmatians as they were already present west of the Don River around this time (Istvánovits & Kulcsár, 2017). Furthermore, the very early Iron Age individual LMO-8 also shares the Hg R-Z93 further reinforcing his connection to the Steppe Sarmatians. Besides him another Early Iron Age individual MJ-31 (Ukraine_Cimmerians_o1, 1284-1055 calBCE) exhibits a similar haplogroup R-Z645 which is on the same branch as R-Z93 but higher.

While the most prevalent Y-Hg of the Steppe Sarmatians is the R1a~ subclade R-Z93 much more R1 haplotype present themselves from higher branches among them. We think it's important to emphasize that at least some of these haplotypes may very well belong under R-Z93 as many of the available genomes exhibit quite poor coverage from capture data and in most cases this prevents the terminal classification of their haplotypes.

The most prevalent haplogroup among our samples was the R1a~ subclade R1a1a1b2a~ (R-Z94) which constituted 4/9 of the Romanian Sarmatian males, 17/67 of the Sarmatian-period males and 6/14 of the Hun-period males. This could be further separated to the subbranches R1a1a1b2a2a~ (R-Z2125) chiefly present among the Romanian Sarmatians and the Hun-period individuals and R1a1a1b2a2b~ (R-Z2122) most prevalent among the bulk of the Carpathian Basin Sarmatians. The subclade R-Z2125 also appears in several Steppe Sarmatians from Kazakhstan (I11540) and Russia (DA144), in the Early Iron Age individual from Romania (LMO-8), in the East Asian immigrant Hun MSG-1, in several Avar-period individuals from Szarvas – Grexatéglagyár and in the Avar-period elite outlier DK-701. The other subclade R-Z2122 represent itself in the Middle and Late Sarmatian-period cemeteries of Madaras – Halmok, Apátfalva - Nagyút dűlő, Apc - Farkas-major and Kiskundorozsma – Subasa among others which were also related according to the IBD analysis. Some further individuals include multiple Steppe Sarmatians from Russia (DA134, DA145 and Pr4) the Conquest-period elite and East Asian outlier K1-3286, the single Conquest-period individual from Árkus (ARK-14) and VPB-279 classified as a Carolingian according to their archaeological findings.

Finally, the influx of Far Eastern Y-Hgs is observed during the Avar Period, with similar changes noted during the Conquest Period as previously reported (Csáky et al., 2020; Neparáczki et al., 2019).

IV.6.2 Persistence of the maternal haplogroup distribution

The mitochondrial data shows (**Figure 23**) that after the Neolithic transition, a relatively stable Mt-Hg frequency pattern persisted in the Carpathian Basin. This is in stark contrast compared to the Y-Hg frequency patterns where a major population turnover event seems to be present also at the onset of the Bronze Age. A similar pattern is also apparent in the Sarmatian Period as Y-chromosome data indicates at least a moderate turnover with the sudden appearance of the R1a subclade Z93. This emergence of Asian haplotypes however is not reflected in the Mt-Hg record, as a major influx of Asian haplotypes seemingly only appears in the Avar Period and persists temporarily as it already shows the signs of decline in the following Conquest Period.

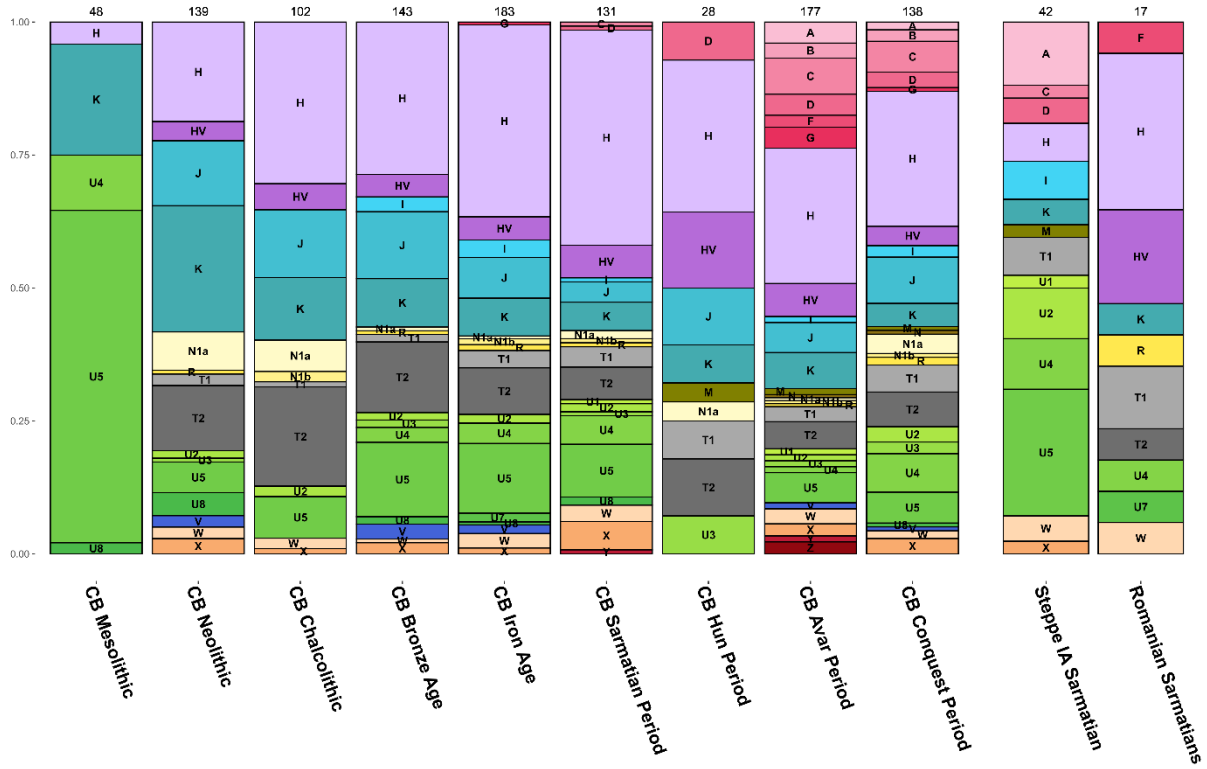


Figure 23: Mitochondrial haplogroup distribution across successive periods in Carpathian Basin populations. The list of individuals contained in each period can be found in **Table E4**.

Asian maternal haplotypes are present among the Steppe Sarmatians in a moderate ratio, this however is already much higher than can be observed among the Romanian Sarmatians. These results posit the possibility that the westward migration of the Sarmatians were mainly driven by male participants, as their mark can be clearly distinguished in the Periods following the Iron Age of the Carpathian Basin, while maternal haplogroup patterns does not change dramatically until the onset of the Avar Period.

Another possibility has to be also considered as the Steppe Sarmatians also exhibit several haplogroups that were prevalent in the Carpathian Basin before their arrival (namely H, K, T1, U4, U5, W and X). This can suggest that a balanced population may have arrived, however their "local-like" haplogroup distribution seamlessly dissolved among the autochthonous population. This however would seem to suggest a relatively small population where drift could significantly alter the descending Hg distributions as Hgs A, C and D are only marginally present among the Carpathian Basin Sarmatians while constituting ~20% of the Hgs for the Steppe Sarmatians.

V DISCUSSION

V.1.1 Limitations

Most of our samples were associated with at least some archaeological context which aided our accurate classification effort. For a detailed description of the archaeological material see the supplement of Schütz et al. (2025). However, due to the frequent disturbance of the Sarmatian-period graves as well as the general poor preservation of their anthropological material, we refrained from drawing inferences based on archaeology in order to avoid erroneous conclusions.

Regarding the genetic results, we need to emphasize the lack of comparative samples from the Carpathian foothills in Romania, dating to the period preceding the 1st century CE arrival of the Romanian Sarmatians. Additionally, a significant portion of the available Iron Age genomes from neighbouring regions (Moldova, Ukraine) were either not sequenced using the shotgun method or have low coverage, making them unsuitable for imputation and IBD analysis in our framework. To some extent, this issue also applies to the Great Hungarian Plain, where Iron Age samples remain scarce. Moreover, our sampling of Sarmatian-era individuals was restricted to the Carpathian Basin and present-day Romania, thus samples from the northwestern end of the Steppe (e.g.: from southern Poland) are missing from our database, which limits our ability to draw definitive conclusions.

These constraints hinder a detailed understanding of population dynamics both beyond the Carpathians and within the Carpathian Basin in the period preceding the Sarmatian migration into the region. Furthermore, the high genetic similarity among European genomes from different sub-regions during this period hinders the precise identification of “local” genome components, using statistical methods such as F statistics and qpAdm.

V.1.2 Relation to Steppe Sarmatians

To study the origin and genetic relations of Carpathian Basin Sarmatians, we have compiled the most representative database of this era and region to date. We have shown that the Carpathian Basin Sarmatians differed significantly from Steppe Sarmatians and more closely resembled contemporary populations from their surroundings, except for their small but significant ANA component (**IV.2.2**). In contrast, most Sarmatians from outside the Carpathians in Romania were more similar to the Steppe Sarmatians and appeared to form a genetic link between the two groups.

It is quite striking that the Romanian and Carpathian Basin Sarmatians shared negligible IBDs with the geographically closest and immediately preceding Western Scythians from Hungary, Moldova, and Ukraine. Instead, their immediate source populations, the Steppe Sarmatians, evidently derived from the more remote Ural and Kazakh regions (**Figure 11** and **IV.5.5.1**).

We demonstrated that most studied Sarmatians require a Steppe Sarmatian source in their genome modelling (**IV.4.2.1** and **Table E2**), with this component gradually diminishing over time. This pattern suggest a possible founder effect from limited migration, indicating that Romanian Sarmatians are probably close descendants of Steppe Sarmatians who mixed with the local population after migrating to the Carpathian Basin. IBD data supports this, showing strong connections between Steppe, Romanian, and Carpathian Basin Sarmatians (**IV.5.4.1, Figure 21** and **IV.5.5.2).** Notably, the FKD-150 (Füzesabony – Kastélydűlő, no. 150) female from the HUN_SARM_EP group shares 4 IBD segments totalling 48.5 cM with the DA139 Steppe Sarmatian female from the Pontic Steppe. DA139, in turn, shares 88 cM IBD with the chy001 Uralic Sarmatian and 64 cM with the POG-10 (Pogorăști, M.10) Romanian Sarmatian.

During the Sarmatian Period, the male lineage composition in the Carpathian Basin underwent significant changes, highlighted by the rapid spread of R1a lineages (**IV.6.1**). Notably, the Asian R1a-Z93 subclade gained prominence, clearly originating from the Steppe and Romanian Sarmatians, where this lineage is particularly common. It is worth noting that the available Steppe Sarmatian genomes generally have low coverage, so a large proportion of those classified under the broader R1a1a1 haplogroup possibly belong to the R1a1a1b2-Z93 subclade. In contrast, the maternal lineages did not undergo significant changes, which suggests that the westward migration of the Sarmatians may have been primarily driven by male participants. Nevertheless, it is noteworthy that in the Early Sarmatian-period cemeteries (HUN_SARM_EP), female burials, such as those in the 'Golden Horizon' graves, were especially prominent.

The pronounced intragroup connectedness observed in the Steppe Sarmatians and other recently arriving groups from the steppe (HUN_AVAR_AC and HUN_CONQ_AC) suggests that the nomadic lifestyle could pose as a primary factor responsible for the exceptionally high levels of intragroup IBD sharing (**IV.5.2**). As the Sarmatians settled in the Carpathian Basin they transitioned to a more sedentary lifestyle, with agriculture becoming predominant (Istvánovits & Kulcsár, 2017). The subsequent decrease in intragroup connectedness in later Sarmatian groups is likely attributed to this lifestyle shift and the increasing population size.

V.1.3 Multiple new migration waves

Based on changes observed in the Sarmatian archaeological material, archaeologists hypothesize multiple waves of Sarmatian migration, suggesting the arrival of new populations during the late 2nd and 4th centuries. Our findings support this, indicating that new groups likely arrived during both the Early-Middle and Late Sarmatian Periods, roughly aligning with the archaeological timeline.

In the Early-Middle Period, the populations of the Hódmezővásárhely – Fehértó (HVF) and Makó - Igási Járandó (MIJ) cemeteries show significant differences from both the Sarmatians and earlier local populations, and they share substantial IBD connections with Avar and Conquest-period populations (**Figure 13** and **IV.5.2**).

The Hódmezővásárhely – Fehértó (HVF) individuals show a genetic shift towards northern European populations, with qpAdm analysis confirming the presence of northern European genome types in this cemetery (**Table E2**). For example, HVF-4 and HVF-21 can be exclusively modelled from Scandinavian genomes, while HVF-8 forms a clade with the Poland Wielbark population. The local component of the remaining HVF individuals were typically modelled from Germany_EMedieval_Alemanic_SEurope (O'Sullivan et al., 2018).

The Makó - Igási Járandó (MIJ) individuals, on the other hand, appear to carry northern European genomes admixed with East Asian ones, with most of them significantly shifted towards Asia in PCA (**Figures 4** and **5**). Their local components were typically modelled from Germany_EMedieval_Alemanic_SEurope, while MIJ-1 and MIJ-3 also carry 25% and 20% Xiongnu/Hun-Elite ancestry, respectively. Additionally, MIJ-7 and HVF-2, women with Sarmatian genetic affinity, are closely related, sharing 6 IBD segments with a total length of 142 cM.

These findings suggest that during the Early-Middle Sarmatian Period, there may have been two distinct migration waves: one from northwestern Europe, possibly related to the Germanic tribes of the Marcomannic Wars, and another from the Eastern Steppe, consisting of a population of East Asian origin distinct from the Sarmatians (see also **IV.4.1**).

A likely second wave of migration detected in the Late Sarmatian Period is evident from individuals in the Óföldeák - Ürmös (OFU), Tiszadob – Sziget (TD), Csanádpalota – Országhatár (CSO), and Szihalom - Pamlényi tábla (SPT) cemeteries (**Figure 15**). These individuals have sparse IBD connections with the Sarmatians but show significant ties to the Avar Period. On PCA, they align with the local European population (**Figure 5**), with three

individuals showing a clear shift towards modern Southern Europeans. In qpAdm analysis, all these individuals exhibited the most significant affinity to sources related to the Roman Empire (e.g., Germany_Roman.SG, Italy_IA_Republic.SG, Austria_Ovilava_Roman.SG, Italy_Imperial.SG), with some local and Sarmatian admixture (**Table E2**). Therefore, in the Late Sarmatian Period, the new migration likely came from neighbouring Roman provinces rather than from the steppes.

It is important to note that, genetically, we cannot detect the possible new influx of groups with a similar composition to the first Sarmatian wave, especially if these migrations followed a stepping-stone pattern. For instance, while archaeological data suggests possible elite migrations from the East between the late 2nd century and early 3rd century (Kulcsár, 1998b), these have not been detected genetically.

All analyses identified 5 outlier individuals with elevated ANA ancestry from the Sarmatian Period, which cannot be attributed to Steppe Sarmatians. These individuals were excavated from the Makó - Igási Járandó (MIJ), Madaras – Halmok (MDH) and Nagykálló - Kis Ludas-tó dűlő (NKL) cemeteries. The two individuals from the MIJ site have already been discussed above. MDH-209, dating to the Middle-Late Sarmatian Period, shares IBD segments with multiple Avar-period samples, as well as with individuals from the Early Sarmatian, Hun, and Roman Periods (**Figure 19**).

The Nagykálló - Kis Ludas-tó dűlő (NKL) cemetery is divided archaeologically into two sections, one from the Sarmatian Period and the other from the Hun Period. However, all four Sarmatian NKL samples were grouped into the uncertain category (HUN_SARM_UP) because their radiocarbon dates were spread across an unrealistically wide time range (**Table S3**). Despite this, the eastern components of the NKL-7, NKL-135, and NKL-157 outliers are consistently modelled from Xiongnu/Hun elite ancestry, and they share IBD segments almost exclusively with Avar samples, including Avar elites (**Table E2**). These findings strongly suggest that these individuals are most likely associated with migrations during the Hun Period, though they may have arrived somewhat earlier (see also **IV.2.1**).

The Hun-period samples are distinctly separated into two IBD-sharing clusters (**Figure 19**). Nearly all of the newly sequenced HUN_HUN genomes carry local genome types and are associated with the Sarmatian cluster, including Steppe Sarmatians, with only marginal connections to Roman, Avar, and Conquest-period individuals. In contrast, the previously published Hun-era genomes (Maróti et al., 2022) contain significant Asian components and align with the Avar cluster, including a few Conquest-period elites. These results clearly

indicate that during the Hun Period, most of the Carpathian Basin population represented the existing local population, while new Hun-era immigrants with Asian roots were in the minority. This is entirely consistent with historical data, which indicates that Sarmatian cemeteries were used until the early 5th century, and settlements until the mid-5th century.

Particularly noteworthy is the genome of the ASZK-1 (Árpás - Dombiföld, Szérűskert, no. 1) individual from the Hun Period, which forms a clade with Steppe Sarmatians in most qpAdm models, while also exhibiting some East Asian admixture in other valid models (Maróti et al., 2022). This solitary and rich Hun burial is well-dated and bears numerous parallels to similar finds in the Kazakh Steppe. The burial customs and the entirety of the findings suggest an Eastern individual from an Eastern environment, making it likely that this individual arrived with the Huns. (Tomka, 2001). This genome suggests that the descendants of the Steppe Sarmatians were also present among the incoming Huns.

The IBD connections separate the populations of the three successive migration waves (Sarmatian, Avar and Hungarian Conquest) into three distinct clusters (**Figure 12**), suggesting that the three migration waves are largely associated with different populations.

V.1.4 Iron age samples

The two Early Iron Age samples LMO-8 (Lișcoteanca - Movila Olarului, M.8) and RAM-7 (Râmnicelu, 1969, M.7) from the Carpathian foothills were contemporary neighbours of the European Scythians, radiocarbon dated to 2,6-2,7 kyears BP and predating the first Steppe Sarmatians (**Table S3**). The LMO-8 male had a Scythian-style bronze arrowhead within his ribcage, which may have caused his death, or perhaps he wore it as a necklace (for further details on the specifics of the burials see Schütz et al., 2025). Surprisingly, despite the large temporal and geographical distances, the ADMIXTURE patterns of LMO-8 and RAM-7 were identical to those of the Steppe Sarmatians (**Figure 6**). In qpAdm models, RAM-7 nearly formed a clade with the Steppe Sarmatians, while LMO-8 appeared to be approximately 75-90% Steppe Sarmatian with about 10-25% East or Central Asian admixture (**Table E2**). Additionally, the R1a1a1b2a2a~ (R1a-Z2124) Y-chromosomal haplogroup of LMO-8 aligns with the typical haplogroups found among the Steppe Sarmatians (RAM-7 being female).

Their potential connection to the Steppe Sarmatians is further supported by IBD data. LMO-8 shares IBDs with three Steppe Sarmatians, including one connection that is 22.5 cM long (see also **IV.5.5.1**). Additionally, LMO-8 shares IBDs with a Central Saka and a Tian Shan Saka, as

well as with FKD-150, an early Carpathian Basin Sarmatian. Similarly, RAM-7 shares IBD with two early Sarmatians from the Ural region and with a Scythian from Moldova.

These data demonsrate that the two Iron Age individuals were genetically very similar to the later Sarmatians. This suggests that migrations from the Ural region westward may have already occurred during the Early Iron Age, at least sporadically. It is worth noting that the age of these two individuals is much closer to the European appearance of the Cimmerians, and one well-covered genome identified as Cimmerian (MDA_IA_CIMM, **Table S5**) does indeed show a similar ADMIXTURE pattern (**Table E1**). Thus, it cannot be ruled out that their appearance may be connected to the “Cimmerian” migrations.

VI SUMMARY

VI.1 Summary of the thesis work

The Sarmatians were a group of nomadic people who likely originated from the southern Ural region during the 4th and 2nd centuries BCE. In the subsequent centuries, they gradually expanded into the Pontic Steppe territories, displacing the culturally related Scythians (Koryakova, 2018; Mordvintseva, 2013). During the Iron Age, they established the first significant political formations in the area between the Don, Volga, North Caucasus, and Ural Mountains.

By the first century CE, Sarmatian groups had settled in the area between the eastern foothills of the Carpathians and the Lower Danube region (modern Romania; P. Kovács, 2023). In the early decades CE – according to historical sources – the first Sarmatian tribes, known as the Iazyges, entered the Carpathian Basin, occupying the northern and central areas of the Danube-Tisza interfluve. They then gradually expanded into the Trans-Tisza region, eventually occupying the entire Great Hungarian Plain, and likely extending their rule over the local Celtic and Scythian groups.

Their complicated relationship with the neighbouring Roman Provinces as well as with different Germanic tribes (Quadi, Marcomanni, Vandals) significantly influenced their material culture, particularly in the surrounding border areas (Bene et al., 2016; H. Vaday & Horváth, 2005; Istvánovits & Kulcsár, 2000, 2003, 2020; P. Kovács, 2009; Vaday, 1988, 2001). The dense settlement network in the Carpathian Basin within a century of their arrival indicates that the nomadic herders also adopted farming and achieved a large population size (Istvánovits & Kulcsár, 2015; Masek, 2021; Pető et al., 2017). Despite this, many steppe traditions persisted in daily life, culture, and warfare. Sarmatians in the eastern Carpathian Basin maintained close contact with other Sarmatian groups in the steppe, and archaeological finds show several instances of eastern groups moving in during the late 2nd and 4th centuries (Istvánovits & Kulcsár, 2017; Khrapunov, 2001; Kulcsár, 1998a).

Intriguingly, this once-dominant people, who ruled over a vast region and significantly influenced the ancient and early medieval world (military innovations, relations with the Roman Empire, and even ties to the Arthurian legend) are not claimed as ancestors by any modern European state-forming nations and remain a group of ancient, now forgotten people (Istvánovits & Kulcsár, 2017).

Between the late 4th and mid-5th centuries, a major migration initiated by the Huns brought diverse communities, including Eastern Germanic tribes, Huns, and other eastern Sarmatian groups, into the Great Hungarian Plain (Soós, 2019b; Tejral, 2011). After the Hunnic empire moved its centre to the Carpathian Basin, many Sarmatians remained in their original homeland. Their cemeteries were used until the early 5th century, and their settlements continued until the mid-5th century collapse of the Hun Empire (H. Vaday, 1994; Masek, 2021; Tejral, 2011). According to written sources, the Sarmatians may have maintained an independent political organization until the 470s (P. Kovács, 2023). After the fall of the Hun Empire however, they were assimilated into the population of the Gepid Kingdom (Kiss, 2015; Kiss P., 2021).

To date, 45 published Sarmatian genomes, are available across 7 different studies, from the Ural region and the Central Steppe (Damgaard et al., 2018; Gnecchi-Ruscone et al., 2021; Järve et al., 2019; Krzewińska et al., 2018; Narasimhan et al., 2019; Unterländer et al., 2017; Veeramah et al., 2018). Among these studies, only two articles provide a detailed discussion of the Sarmatians in context (Gnecchi-Ruscone et al., 2021; Krzewińska et al., 2018) while the others address the topic only marginally. Key characteristics identified in the Uralic Sarmatians and Easten Steppe Sarmatians (referred to as Steppe Sarmatians) include:

a) Their genomes exhibit the admixture of three main ancestral components: 70% Steppe Middle-Late Bronze Age (steppe_MLBA), 18% Bactria–Margiana Archaeological Complex (BMAC)-related, and 12% Baikal Early Bronze Age (Baikal_EBA)-Khovsgol-related.

b) Their lower Khovsgol-related East Asian component compared to the Eastern Scythians suggests they might have originated from distinct, independent late Bronze Age populations in the Ural area.

c) Despite their extensive geographical distribution and relatively high genetic diversity, they remained genetically very homogeneous for over 500 years.

From the Carpathian Basin 17 Sarmatian-Period individuals were published in (Gnecchi-Ruscone et al., 2022). While detailed analyses were not provided, these genomes display a distinct shift towards European genetic profiles compared to Steppe Sarmatians, raising questions about the potential relationships between these populations.

To clarify the origins and genetic relationships of the Carpathian Basin Sarmatians and to explore their connections to other populations from the Eurasian Steppe, as well as to local groups from preceding and succeeding periods, we sequenced 156 genomes from the Carpathian Basin and surrounding regions, spanning the Sarmatian and Hun Periods. 17

samples were excavated in present day Romania, outside of the Carpathian Basin. We acquired these samples to test the proposed southern migration route for the alleged arrival of the Steppe Sarmatians as well as to study possible genetic links between the Sarmatians residing in the Carpathian Basin and on the Western Steppes. 22 samples were collected from the subsequent Hun Period to investigate demographic changes following the onset of the Great Migration Period and to assess the potential continuity of the Sarmatian-era population. Finally, we incorporated in our data set the above mentioned 17 Sarmatian genomes published in Gnecchi-Ruscone et al. (2022) as well as 9 Hun-period samples published in our previous article (Maróti et al., 2022). This resulted in the most comprehensive currently available genome database that represents the population of the Carpathian Basin between the 1-5th centuries AD.

PCA results showed that the majority of the Carpathian Basin Sarmatians occupy a distinct position compared to the Steppe Sarmatians substantially shifted toward the modern Central European cluster. Despite this, several individuals (especially from the early Sarmatian Period of the Carpathian Basin, as well as from Romania) exhibited clear affinity toward the Steppe Sarmatians. Furthermore, a number of outliers fell outside the Carpathian Basin Sarmatian cluster rather mapping towards modern Northern Europeans or toward present-day East Asians. These cases likely represent migrant groups whose origins were independent of the Steppe Sarmatians.

ADMIXTURE analysis produced similar results, revealing that most of the newly sequenced individuals shared greater genetic similarity with previous and contemporary populations of the Carpathian Basin than with the Steppe Sarmatians. Nevertheless, a small but detectable East Asian–related genetic component was present in the majority of individuals separating them from other populations of the region.

This component was shown to be adequately modelled from Steppe Sarmatian sources using both F4-statistics and qpAdm. Moreover, qpAdm analysis indicated that approximately 2/3 of the individuals included Steppe Sarmatian ancestry in their admixture models, at least as a minor component.

To gain further insight, we investigated potential genealogical connections across an extensive ancient Eurasian whole-genome database using identity-by-descent (IBD) analysis. Initial results showed that the Carpathian Basin Sarmatians – as a group – exhibited the strongest IBD links to Hun, Avar, and Conquest-period individuals from the same region, as well as to Steppe Sarmatians. We continued our investigation on a subset of the original database containing only these groups. The resulting IBD network revealed that the Carpathian Basin Sarmatians, Avar,

and Conquest-period individuals formed three distinct clusters, connected by a sparse but detectable network. In contrast, individuals from the Hun Period did not form a distinct cluster, instead appearing dispersed among individuals assigned to the previous and subsequent periods. Notably, the Steppe Sarmatians displayed strong affinity to the Carpathian Basin Sarmatians, occupying a central position within the Sarmatian cluster.

Further IBD count-based analyses showed that the Steppe Sarmatians indeed share a substantial amount of IBDs with the different Carpathian Basin Sarmatian groups and this connection seemed to gradually decline throughout the progressive periods. This indicated a possible founding effect originating from the Steppe Sarmatians with a continual chain of generational transmission. Possible migration events could be identified in the Early-Middle and Late Sarmatian Periods pointing toward Northern Europe, the Central Steppe and the Roman Empire as likely sources. Notably, the Hun-period group exhibited the lowest rate of intragroup connectedness, while their intergroup connection level – compared to this – remained much higher. This pattern supports the hypothesis that the Hun-period assemblage is somewhat artificial: while some individuals clearly reflect new migration events, others appear to represent the continuation of the previous Sarmatian population. Finally, both the Sarmatian and Hun-period group projected a low but detectable ratio of IBD connections toward the much later Avar and Conquest-period population, indicating at least partial survival across these transitions.

By analysing close genealogical connections, we identified an extended family of Steppe Sarmatian individuals that spans the full duration of Sarmatian occupation across the Steppe region, both spatially and temporally. Remarkably, this family also included several individuals sequenced in our study, among them 3 Sarmatians from Romania (POG-10, RLS-1, TAR-196) and 1 Early Sarmatian from the northern edge of the Great Hungarian Plain (FKD-150). This finding further reinforced the apparent deep-rooted connection among the different Sarmatian groups across the Western Steppe.

Haploid data analysis showed a very apparent turnover in the Y-chromosome haplogroup distribution of the Sarmatian Era population of the Carpathian Basin. The characteristic Central Asian haplogroup R1a-Z94, which had been only sporadically present in earlier periods, emerged suddenly in this era. It was also the most common Y-haplogroup among Steppe Sarmatian males and in our Romanian Sarmatian samples, highlighting a strong unifying paternal signal across all Sarmatian groups. In contrast, the maternal haplogroup distribution showed less coherence. While many maternal lineages found among the Steppe Sarmatians are

also typical of Central Europe, a notable proportion of Asian-derived mitochondrial haplogroups was present among them. These however were largely absent in both the Carpathian Basin and Romanian Sarmatian samples. A significant shift in maternal lineages only became evident after the Sarmatian Period and particularly during the much later Avar Period. Taken together, these findings suggest that the Sarmatian migration was predominantly male driven.

In summary, we show that the Carpathian Basin Sarmatians descended from Steppe Sarmatians originating in the Ural and Kazakhstan regions, with Romanian Sarmatians serving as a possible genetic bridge between the two groups. However, the steppe-derived ancestry observed in the Carpathian Basin appears significantly diluted, likely due to substantial local admixture or the migration of groups that were already genetically admixed prior to their arrival. We also identify two previously unknown migration waves during the Sarmatian era and a notable continuity of the Sarmatian population into the Hunnic period, despite a smaller influx of Asian-origin individuals. These results shed new light on Sarmatian migrations and the genetic history of a key population neighbouring the Roman Empire.

VI.2 Magyar nyelvű összefoglaló

A szarmaták nomád törzsek csoportjainak összefoglaló neve. Eredetük minden valószínűség szerint a Dél-Urál vidékére vezethető vissza megközelítőleg az i.e. 4-2 századba. Az i.e. 3-2 századtól nyugat felé kezdtek terjeszkedni és sorozatos konfliktusok nyomán fokozatosan kiterjesztették hatalmukat a korábban szkíta uralom alatt álló területekre (Koryakova, 2018; Mordvintseva, 2013).

A történeti források szerint az i.sz. 1. században egyes szarmata törzsek (elsősorban a Jazigok) a Kárpát-medencébe is bevándoroltak (P. Kovács, 2023). Először a Duna-Tisza közén telepedtek le, de később az egész Alföld területét uralmuk alá hajtották. A Római Birodalommal, valamit a szomszédos germán és kelta törzsekkel fennálló komplex kapcsolati rendszerük jelentős hatást gyakorolt az anyagi kultúrájukra (Bene et al., 2016; H. Vaday & Horváth, 2005; Istvánovits & Kulcsár, 2000, 2003, 2020; P. Kovács, 2009; Vaday, 1988, 2001). Bár kultúrájukban bizonyos steppei hagyományok sokáig felfedezhetők maradtak, a korszakból feltárt sűrű település hálózat arra utal, hogy a Kárpát-medencébe érkező szarmaták letelepedtek és nagyrészt növénytermesztő gazdálkodásra tértek át, mellyel együtt a populáció méretük is jelentősen megnőtt (Istvánovits & Kulcsár, 2015; Masek, 2021; Pető et al., 2017).

A 4-5. században megkezdődő nagy népvándorlási hullámok változatos eredetű csoportokat sodortak a régióba, melyek folytonos veszélyt jelentettek a szarmatákra (Soós, 2019b; Tejral, 2011). A népmozgásokba maguk is belekeveredtek, egyes csoportjaik a Kárpát-medencét elhagyva a Római Birodalom területére vándoroltak (Istvánovits & Kulcsár, 2017). Politikai önállóságuknak a hunok érkezése vetett véget, akik az 5. század első felében a régióba helyezték át központjukat. Ennek ellenére sok szarmata település a régészeti adatok alapján a hun uralom végéig (450-es évek) használatban maradt (H. Vaday, 1994; Masek, 2021; Tejral, 2011). A Hun Birodalom bukását követően, a történeti forrásokban még találhatók utalások a szarmaták Kárpát-medencei tevékenységeiről egészen az 5. század végéig, ám ezután a gepidák politikai uralma alá került a régió, akik valószínűleg asszimilálták a megmaradt szarmatákat (Kiss, 2015; Kiss P., 2021).

A majd 400 éves szarmata uralom minden valószínűség szerint jelentős hatást gyakorolt a Kárpát-medence népességtörténetére. Ennek ellenére ez az egykoron jelentőségteljes nép mára jórészt elfeledetté vált és egyik újkori állam alkotó nemzet sem vállalt velük ősiséget (Istvánovits & Kulcsár, 2017).

A szakirodalomban eddig 7 publikációban összesen 45 steppei szarmatákhoz köthető genomot közöltek (Damgaard et al., 2018; Gnecchi-Ruscone et al., 2021; Järve et al., 2019; Krzewińska et al., 2018; Narasimhan et al., 2019; Unterländer et al., 2017; Veeramah et al., 2018). Ezen munkák nagy része első sorban a szkíták eredetével foglalkozott, és a szarmatákat 2 munkától eltekintve (Gnecchi-Ruscone et al., 2021; Krzewińska et al., 2018) nem tárgyalták részletesen. A közelmúltban további 17 szarmatákhoz köthető genomot közöltek (Gnecchi-Ruscone et al., 2022), ezúttal a Kárpát-medencéből. Bár ezek a genomok kimaradtak a részletes analízisekből, az általános genetikai összetételük jelentős eltérést mutatott a steppén feltárt szarmatákhoz képest, sokkal inkább a helyi, közép-európai genomokhoz hasonlítottak.

Mivel a Kárpát-medencei szarmaták pontos eredetét máig homály fedi, ezért célul tűztük ki a Kárpát-medence szarmata kori népességének populáció genetikai vizsgálatát. A kutatásunk során 156 szarmata és hun kori egyénből állítottunk elő teljes genom szekvenciákat. 17 egyén a Kárpát-medencén kívülről, a mai Románia területéről származott. Ezeket azért vontuk be a vizsgálatokba, hogy segítsenek eldönteni a történeti forrásokban említett déli irányú bevándorlás kérdését. 22 egyént a szarmata kort követő hun korszakból mintáztunk. Segítségükkel a népvándorlás kori demográfiai változásokat kívántuk vizsgálni, ezen felül pedig a szarmaták esetleges tovább éléséről is értékes információt szolgáltattak. Bevontuk még az adatbázisunkba a korábban említett 17 Kárpát-medencei szarmata egyént, melyet Gnecchi-

Ruscone és munkatársai közöltek (Gnecchi-Ruscone et al., 2022), valamint 9 hun kori egyént a korábbi publikációnkból (Maróti et al., 2022). Összesítve 182 egyén teljes genom adatait gyűjtöttük össze, mely az i.sz. 1-5 század közötti Kárpát-medence legteljesebb populáció genetikai adatbázisának tekinthető.

A genomok általános struktúráját a hipotézis független Főkomponens Analízis (PCA) és ADMIXTURE módszerek segítségével vizsgáltuk (Alexander et al., 2009; Patterson et al., 2006). Ezek egybevágó eredménye azt mutatta, hogy a Kárpát-medencében feltárt szarmaták döntő többsége – genetikai összetételük alapján – egyértelműen elkülönült a dél Urál régióban és a kazak steppén feltárt szarmatáktól (továbbiakban steppei szarmaták) és sokkal inkább a helyi európai populációkhoz hasonlítottak. Ennek ellenére több egyént is találtunk, akik erős affinitást mutattak a steppei szarmaták irányába, különösen a Romániában feltárt és a korai Kárpát-medencei szarmaták között. Ezeken felül az ADMIXTURE analízis a legtöbb egyénben egy alacsony, de szignifikáns arányú kelet ázsiai eredetű genom komponens jelenlétét mutatta, amely elkülönítette a szarmatáinkat a korszakból származó egyéb csoportoktól.

A továbbiakban ezt az ázsiai affinitást vizsgáltuk F4-statisztikák és qpAdm analízis segítségével. A steppei szarmaták mindkét analízisben megfelelőnek bizonyultak az ázsiai összetevők modellezésére. A qpAdm modellek a vizsgált egyének ~2/3-ában jeleztek steppei szarmata forrást.

A genom statisztikai analízist a szakirodalomban elérhető legnagyobb felbontású módszerrel, az IBD (Identity-by-descent) fragmens analízissel egészítettük ki. A módszer segítségével két emberben megegyező DNS szakaszok azonosíthatók, melyek kellően hosszúak ahhoz, hogy bizonyosan egy közös őstől származzanak. Ily módon ez a módszer képes valós genealógiai kapcsolatok kimutatására, akár több száz éves időtávlatban is.

A mintáink IBD kapcsolati hálóját egy nagyméretű eurázsiai genomokat tartalmazó adatbázison vizsgáltuk. Csoport szinten, a Kárpát-medencei szarmaták a legtöbb kapcsolatot a Kárpát-medencei hun, avar és honfoglalás kori egyénekkel, valamint a steppei szarmatákkal mutatták. A további vizsgálatokat egy csökkentett adatszetten végeztük, melyben már csak az erős kapcsolatot mutató csoportok szerepeltek. Az így előállított IBD hálózatban a Kárpát-medencei szarmata, avar és honfoglalás kori egyének 3 többé-kevésbé jól elkülönülő csoportot alkottak, melyeket ritkább kapcsolati háló kötött össze. A hun korszakból származó egyének ezzel ellentétben, nem alkottak elkülönült csoportot, hanem a korábbi és későbbi korszakokból származó egyének között helyezkedtek el. A steppei szarmaták idő- és térbeli távolságuk ellenére igen sűrű kapcsolati hálózattal rendelkeztek mind egymással mind a többi régió

szarmatáival, és a Kárpát-medencei szarmaták által meghatározott „felhő” közepén helyezkedtek el.

A csoportok részletesebb kapcsolatait normalizált IBD szám alapú analízissel vizsgáltuk tovább. Ennek során a csoportok között megfigyelt kapcsolatok számát a lehetséges összes kapcsolat számával normalizáltuk. A steppei szarmaták ez alapján is nagyszámú kapcsolatot mutattak a Kárpát-medencei szarmatákkal, mely fokozatosan csökkent az időben egymást követő korszakok során. Ez a mintázat a steppei szarmaták alapító hatására utalhat, melynek során egy bevándorolt csoport öröksége egy folytonos leszármazási láncon keresztül folyamatosan hígult egészen a szarmata kor végéig. Az egymást követő szarmata korszakok IBD kapcsolatainak eloszlása alapján 2 lehetséges további migrációs eseményt is azonosítottunk a kora-közép és a késő szarmata korokban. Az újonnan érkezettek Észak-Európával, a közép-ázsiai steppével, valamint a Római Birodalom lakosságával mutattak kapcsolatot. A csoporton belüli leggyengébb kapcsolati hálót a hun korszakból származó egyének mutatták, ezzel szemben a csoportok közötti kapcsolataik száma jóval magasabbnak mutatkozott. Bár a korábbi publikációnkban (Maróti et al., 2022) több Kelet-Ázsiából származtatható egyént is azonosítottunk a korszakból, az újonnan szevenált 22 hun kori minta között – 1 kivételével (VIG-94-1) - nem szerepelt több ázsiai bevándorló. Genetikai összetételük sokkal inkább a szarmata kor népességére hasonlított, és az IBD analízisben is ezzel a csoporttal mutatták a legtöbb kapcsolatot. Végezetül mind a szarmata, mind a hun kori csoportoknál megfigyelhetők voltak a későbbi avar és honfoglalás kori népességgel mutatott alacsony, de jól detektálható IBD kapcsolatok, melyek a szarmata népesség legalább részleges tovább élesére utalnak.

A közeli genealógiai kapcsolatok vizsgálata során azonosítottunk egy kiterjedt steppei szarmata családot, mely a teljes szarmata uralom időszakát lefedte. A család legkorábbi tagjait 3 Dél-Urál régióból származó korai szarmata egyén alkotta (i.e. ~ 400), míg a későbbi egyének között szerepelt 3 általunk szekvenált szarmata Romániából és 1 korai szarmata a Kárpát-medencéből (Füzesabony – Kastélydűlő, 150-es sírszám). Ezen rokoni kötelékek egyértelműen bizonyítják a Kárpát-medence és az Urál vidék szarmatáinak közvetlen kapcsolatát, de a steppe teljes szarmata népességének összefonódására is rámutatnak.

Az apai és anyai vonalak vizsgálata is fontos új eredményeket adott. A jellegzetes közép-ázsiai elterjedésű R1a-Z94, Y-kromoszómás haplocsoport a bronzkor óta csak sporadikusan fordult elő a Kárpát-medencében, ezzel szemben a szarmata korban ez vált a leggyakoribb apai vonallá. Mind a steppei szarmaták, mind a Romániában feltárt szarmaták között a R1a-Z94 haplocsoport

volt a leggyakoribb, ami újabb bizonyíték az uráli, romániai és Kárpát-medencei szarmata népesség mélyen gyökerező kapcsolatára. Ezzel ellentétben az anyai vonalak vizsgálata nem mutatott jelentős változást. Bár a steppei szarmaták mitokondriális haplocsoport eloszlása hasonlónak mutatkozott a Kárpát-medencei népességek haplocsoport eloszlásához, de emellett jelentős arányú ázsiai eredetű anyai vonalak (A, C, D) is megtalálhatók voltak közöttük. Ezek a vonalak azonban hiányoztak a Kárpát-medence vaskori népességéből, ahol a neolitikum óta többé-kevésbé stabil haplocsoport eloszlás figyelhető meg, és csak a jóval későbbi avar kortól kezdődően detektálható jelentős ázsiai haplocsoport beáramlás. Ezen eredmények arra utalnak, hogy a szarmaták Kárpát-medencei bevándorlása során többségében férfiak érkezhettek.

Összefoglalásként kimutattuk, hogy a Kárpát-medencében feltárt szarmata kori népesség jelentős része a steppei szarmatáktól származtatható, akik először a mai Románia területére érkeztek, és onnan vonultak tovább az Alföld területére. Útjuk során, és főként a Kárpát-medencében a steppei genetikai örökségük jelentősen lecsökkent, valószínűsíthetően a helyi népekkel való keveredésük következtében. A részletes IBD elemzés révén két további bevándorlási eseményt is azonosítottunk, valamint kimutattuk a Kárpát-medencei szarmaták jelentős mértékű továbbélését a hun korszakban és legalább részleges továbbélését a jóval későbbi avar és honfoglalás korokban. Eredményeink új adatokat szolgáltatnak a Kárpát-medence populáció történetének egy meghatározó szeletéről és egy olyan jelentős népességről, mely napjainkra sajnálatos módon feledésbe merült.

VII ACKNOWLEDGEMENTS

This project was, above all, a collaborative effort, and I am honoured to present the results of the work carried out by many dedicated experts. I am also deeply grateful to those who have supported me throughout my life and made it possible for me to pursue my PhD and complete this thesis. Here, I would like to thank at least some of them for their invaluable contributions.

First and foremost, I would like to express my sincere gratitude to my supervisor, **Dr. Tibor Török**, who has accompanied me throughout nearly the entire course of my professional development. I am profoundly indebted to him for his guidance and support. His personal character and philosophy have exerted a great impact on my growth both on a scientific and individual level.

I would also like to thank **Dr. Zoltán Maróti** for his expert contributions, both generally and specifically to this project. His methodological improvements greatly elevated the scientific value of this research. His unique approach and innovative thinking continue to inspire me.

I am thankful to **Dr. István Raskó**, who closely followed the progress of our group, including my own work. His genuine and memorable advices remain invaluable and always appreciated.

My gratitude extends to my colleague **Dr. Gergely I. Bobek Varga** for his professional support, his humour, and the enjoyable hours spent during trips and conferences. I also thank my friends and colleagues **Dr. Alexandra Gînguță**, **Bence Kovács**, and **Dr. Kitti Maár** for their essential contributions in the molecular biology laboratory, as well as for their personal support and the many memorable moments we shared. I am grateful to **Dr. Emil Nyerki** for his important bioinformatics work on this and previous projects, and to **Dr. Olga Spekker** for both her professional guidance and her personal support. I also extend my thanks to my former colleagues **Petra Kiss** and **Dr. Endre Neparáczki** for their valuable contributions and for the time we spent working together.

Special thanks to **Dr. Attila P. Kiss** and **Dr. Balázs Tihanyi** for their crucial guidance on the historical and archaeological aspects of this research, and for their personal support.

I would like to acknowledge **Botond Bán**, **Gellért Csikós** and **Virág Török** who prepared their BSc theses at our lab and were a source of great inspirations for me, encouraging me to improve while helping them.

My thanks also go to **Dr. Luca Kis**, with whom I had the pleasure of collaborating on the study of non-metric dental traits, a project that significantly contributed to my scientific development.

I am grateful to my friends **Dr. András Kaisinger**, **Balázs Vásárhelyi**, and **Márton Kopasz** for their companionship, support and the many adventures that we embarked upon through the past four years, as well as to **Anna Sára Piros** and **Rebeka Rapavi**, whose friendship I could always rely on, even in difficult times. I also thank **Valentina Nagy** for her enduring personal support throughout my PhD years.

I would like to acknowledge all my colleagues in the **Department of Genetics** (University of Szeged) for their professional assistance and the many enjoyable moments at department gatherings. My sincere thanks go as well to the numerous archaeologists and anthropologists who contributed to this project, their help was invaluable.

Finally, I express my deepest gratitude to my family: my mother **Magdolna Mátyus**, my father **Péte Schütz**, my brother **Brúnó Schütz**, and my grandparents **Piroska Pintér Burus**, **János Schütz**, and **Mária Istvánné Mátyus**. Without their support, neither this project nor my PhD would have been possible. I thank them for always being there for me and for giving me the opportunity to become the person I am today.

VIII REFERENCES

Adrews, S. (2023). *FastQC: A quality control tool for high throughput sequence data.* (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; Version 0.12.0) [Java].

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., … Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature*, *522*(7555), 167–172. https://doi.org/10.1038/nature14507

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., … 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

Bârcă, V., & Symonenko, O. (2009). *Călăreţii stepelor. Sarmaţii în spaţiul nord-pontic / Horsemen of the steppes. The Sarmatians in the North Pontic region: Vol. The Centre of Roman Military Studies*. Editura Mega.

Bene, Z., Istvánovits, E., & Kulcsár, V. (2016). Some characteristic types of Roman imports in Sarmatian Barbaricum in the Carpathian Basin (caskets decorated with metal mounts, bronze vessels, mirrors). In N. Müller-Scheeßel & H.-U. Voß (Eds.), *Archäologie zwischen Römern und Barbaren: Zur Datierung und Verbreitung römischer Metallarbeiten des 2. Und 3. Jahrhunderts n. Chr. Im Reich und im Barbaricum—Ausgewählte Beispiele (Gefäße, Fibeln, Bestandteile militärischer Ausrüstung, Kleingerät, Münz* (pp. 743–760). Dr. Rudolf Habelt Verlag.

Bierbrauer, V. (1994). Archäologie und Geschichte der Goten vom 1.–7. Jahrhundert. *Frühmittelalterliche Studien*, *28*, 51–171.

Bierbrauer, V. (1999). Die ethnische Interpretation der Sîntana de Mureş–Černjachov-Kultur. In G. Gomolka-Fuchs (Ed.), *Die Sîntana de Mureş–Černjachov-Kultur: Akten des internationalen Kolloquiums in Caputh vom 20. Bis 24. Oktober 1995; [zum Gedenken an Kazimierz Godłowski, 9. 12. 1934 – 9. 7. 1995]* (pp. 211–238). Habelt.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

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. (2020). Genetic insights into the social organisation of the Avar period elite in the 7th century AD Carpathian Basin. *Scientific Reports*, *10*(1), 948. https://doi.org/10.1038/s41598-019-57378-8

Csárdi, G., Nepusz, T., Traag, V., Horvát, S., Zanini, F., Noom, D., & Müller, K. (2024). *igraph: Network Analysis and Visualization in R*. https://doi.org/10.5281/zenodo.7682609

Damgaard, P. D. B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J. V., Pedersen, M. W., Goldberg, A., Usmanova, E., Baimukhanov, N., Loman, V., Hedeager, L., Pedersen, A. G., Nielsen, K., Afanasiev, G., Akmatov, K., Aldashev, A., Alpaslan, A., …

Willerslev, E. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature*, *557*(7705), 369–374. https://doi.org/10.1038/s41586-018-0094-2

De Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J. V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., Zaibert, V., Merz, V., Biddanda, A., Merz, I., Loman, V., Evdokimov, V., Usmanova, E., Hemphill, B., Seguin-Orlando, A., … Willerslev, E. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*, *360*(6396), eaar7711. https://doi.org/10.1126/science.aar7711

Dinnyés, I. (1991). A hévizgyörki szarmata sírok. Sarmatian graves from Hévizgyörk. *Studia Comitatensia*, *22*, 145–201.

Dunnington, D. (2023). *ggspatial: Spatial Data Framework for ggplot2*.

Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, *12*(1), 14683. https://doi.org/10.1038/s41598-022-14395-4

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18. https://doi.org/10.18637/jss.v048.i04

Flegontova, O., Işıldak, U., Yüncü, E., Williams, M. P., Huber, C. D., Kočí, J., Vyazov, L. A., Changmai, P., & Flegontov, P. (2025). Performance of qpAdm-based screens for genetic admixture on graph–shaped histories and stepping stone landscapes. *Genetics*, *230*(1), iyaf047. https://doi.org/10.1093/genetics/iyaf047

Freeman, L., Brimacombe, C. S., & Elhaik, E. (2020). aYChr-DB: A database of ancient human Y haplogroups. *NAR Genomics and Bioinformatics*, *2*(4), lqaa081. https://doi.org/10.1093/nargab/lqaa081

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prüfer, K., Meyer, M., Kelso, J., Reich, D., & Pääbo, S. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, *524*(7564), 216–219. https://doi.org/10.1038/nature14558

Gnecchi-Ruscone, G. A., Khussainova, E., Kahbatkyzy, N., Musralina, L., Spyrou, M. A., Bianco, R. A., Radzeviciute, R., Martins, N. F. G., Freund, C., Iksan, O., Garshin, A., Zhaniyazov, Z., Bekmanov, B., Kitov, E., Samashev, Z., Beisenov, A., Berezina, N., Berezin, Y., Bíró, A. Z., … Krause, J. (2021). Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Science Advances*, *7*(13), eabe4414. https://doi.org/10.1126/sciadv.abe4414

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., … Krause, J. (2022). Ancient genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites. *Cell*, *185*(8), 1402-1413.e21. https://doi.org/10.1016/j.cell.2022.03.007

Godłowski, K. (1970). *The chronology of the late roman and early migration periods in Central Europe*. Univ. Iagellonicae.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-

Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Grumeza, L. (2014). *Sarmatian cemeteries from Banat (late 1 st – early 5 th centuries AD)*. Mega Publishing House.

H. Vaday, A. (1989a). Die sarmatischen Denkmäler der Sarmatenzeit des Komitats Szolnok. Ein Beitrag zur Archäologie und Geschichte des sarmatischen Barbaricums. *Antaeus*, *17–18*.

H. Vaday, A. (1989b). Sarmatisches Männergrab mit Goldfund aus Dunaharaszti. [Szarmata aranyleletes férfisír Dunaharasztiból.]. *Folia Archaeologica*, *40*, 129–136.

H. Vaday, A. (1991). The Dacian question in the Sarmatian Barbaricum. *Antaeus*, *19–20*, 75–83.

H. Vaday, A. (1994). Late Sarmatian graves and their connections within the Great Hungarian Plain [Neskorosarmatské hroby a ich vzt'ahy v rámci Vel'kej uhorskej nížiny.]. *Slovenská Archeológia*, *42*(1), 105–124.

H. Vaday, A. (1996). Roman Period Barbarian settlement at the site of Gyoma 133. In A. H. Vaday (Ed.), *Cultural and landscpae changes in South-East Hungary 2., Prehistoric, Roman Period Barbarian and Late Avar Settlement at Gyoma 133* (pp. 51–305). Archaeolingua.

H. Vaday, A., & Horváth, F. (2005). *Corpus der römischen Funde im europäischen Barbaricum. Ungarn. 1, Komitat Szolnok*. Archäologisches Institut der UAW.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., … Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207–211. https://doi.org/10.1038/nature14317

Harney, É., Cheronet, O., Fernandes, D. M., Sirak, K., Mah, M., Bernardos, R., Adamski, N., Broomandkhoshbacht, N., Callan, K., Lawson, A. M., Oppenheimer, J., Stewardson, K., Zalzala, F., Anders, A., Candilio, F., Constantinescu, M., Coppa, A., Ciobanu, I., Dani, J., … Pinhasi, R. (2021). A minimally destructive protocol for DNA extraction from ancient teeth. *Genome Research*, *31*(3), 472–483. https://doi.org/10.1101/gr.267534.120

Harney, É., Patterson, N., Reich, D., & Wakeley, J. (2021). Assessing the performance of qpAdm: A statistical tool for studying population admixture. *Genetics*, *217*(4), iyaa045. https://doi.org/10.1093/genetics/iyaa045

Hollister, J., Shah, T., Nowosad, J., Robitaille, A. L., Beck, M. W., & Johnson, M. (2023). *elevatr: Access Elevation Data from Various APIs*. https://doi.org/10.5281/zenodo.8335450

Istvánovits, E., & Kulcsár, V. (2000). Iranian-Germanic contacts in the Sarmatian Barbaricum of the Carpathian basin. In M. Mączyńska & T. Grabarczyk (Eds.), *Die spätrömische Kaiserzeit und die frühe Völkerwanderungszeit in Mittel- und Osteuropa* (pp. 237–260). Wydawnictwo Uniwersytetu Łódzkiego.

Istvánovits, E., & Kulcsár, V. (2003). Some traces of Sarmatian–Germanic contacts in the Great Hungarian Plain. In C. v. Carnap-Bornheim (Ed.), *Kontakt – Kooperation – Konflikt. Germanen und Sarmaten zwischen dem 1. Und dem 4. Jahrhundert nach Christus* (pp. 227–238). Wachholtz Verlag.

Istvánovits, E., & Kulcsár, V. (2006). Az első szarmaták az Alföldön. (Gondolatok a Kárpátmedencei jazig foglalásról.) [The first Sarmatians in the Great Hungarian Plain. (Some notes on the Jazygian immigration into the Carpathian Basin.). *A Nyíregyházi Jósa András Múzeum Évkönyve*, *48*, 203–237.

Istvánovits, E., & Kulcsár, V. (2015). Animals of the Sarmatians in the Carpathian Basin: Archaeozoology through the eyes of archaeologists. *MATERIALY PO ARHEOLOGII ISTORII I ETNOGRAFII TAVRII*, *20*, 49–78.

Istvánovits, E., & Kulcsár, V. (2017). *Sarmatians History and Archaeology of a Forgotten People*. Verlag des Römisch-Germanischen Zentralmuseums.

Istvánovits, E., & Kulcsár, V. (2020). Sarmatians on the Borders of the Roman Empire. Steppe Traditions and Imported Cultural Phenomena. *Ancient Civilizations from Scythia to Siberia*, *26*(2), 391–402. https://doi.org/10.1163/15700577-12341381

Istvánovits, E., Kulcsár, V., & Mérai, D. (2013). Roman Age barbarian pottery workshop in the Great Hungarian Plain. In J. Bemmann, M. Hegewisch, M. Meyer, & M. Schmauder (Eds.), *Drehscheibentöpferei im Barbaricum. Technologietransfer und Professionalisierung eines Handwerks am Rande des Römischen Imperiums. Akten der Internationalen Tagung in Bonn vom 11. Bis 14. Juni 2009.* (pp. 355–369). Inst. für Vor- und Frühgeschichtliche Archäologie der Rheinischen Friedrich-Wilhelms-Univ. Bonn.

Istvánovits, E., Lőrinczy, G., & Pintye, G. (2005). A Szegvár–oromdűlői császárkori telep. [Die frühkaiserzeitliche Siedlung von Szegvár–Oromdűlő.]. *A Móra Ferenc Múzeum Évkönyve – Studia Archaeologia*, *11*, 51–114.

Järve, M., Saag, L., Scheib, C. L., Pathak, A. K., Montinaro, F., Pagani, L., Flores, R., Guellil, M., Saag, L., Tambets, K., Kushniarevich, A., Solnik, A., Varul, L., Zadnikov, S., Petrauskas, O., Avramenko, M., Magomedov, B., Didenko, S., Toshev, G., … Villems, R. (2019). Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology*, *29*(14), 2430-2441.e10. https://doi.org/10.1016/j.cub.2019.06.019

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R. L., Gallego Llorente, M., Cassidy, L. M., Gamba, C., Meshveliani, T., Bar-Yosef, O., Müller, W., Belfer-Cohen, A., Matskevich, Z., Jakeli, N., Higham, T. F. G., Currat, M., Lordkipanidze, D., … Bradley, D. G. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications*, *6*(1), 8912. https://doi.org/10.1038/ncomms9912

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Jull, A. J. T., Burr, G. S., & Hodgins, G. W. L. (2013). Radiocarbon dating, reservoir effects, and calibration. *Quaternary International*, *299*, 64–71. https://doi.org/10.1016/j.quaint.2012.10.028

Kapcsos, N. (2022). *A Maros alsó szakasza szerepének elemzése a késő császár- és a hunkorban az írott és régészeti források tükrében* [PhD Thesis]. Debreceni Tudományegyetem.

Khrapunov, I. N. (2001). On the contacts between the populations of the Crimea and the Carpathian Basin in the Late Roman Period. In E. Istvánovits & V. Kulcsár (Eds.), *International connections of the Barbarians of the Carpathian Basin in the 1st-5th centuries A. D. Proceedings of the International Conference held in 1999 in Aszód and Nyíregyháza.* (pp. 267–274). Jósa András Múzeum.

Kiss, A. P. (2015). *'...Ut strenui viri...' A gepidák Kárpát-medencei története* [Doctoral dissertation, University of Szeged]. https://doi.org/10.14232/phd.2531

Kiss P., A. (2021). Which came first, the chicken or the egg? The ethnic interpretations of the hoards of Șimleu Silvaniei / Szilágysomlyó: A case study in mixed argumentation. In Z. Rácz & G. Szenthe

(Eds.), *Attila's Europe? Structural transformation and strategies of success in the European Hun period.* (pp. 477–500). Hungarian National Museum, Eötvös Lóránd University.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Körösfői, Z. (2024). *Marosszentanna Sântana de Mureș,. Late imperial cemetery along the Mureş*. Jósa András Múzeum.

Koryakova, L. (2018). Europe to Asia. In C. Haselgrove, K. Rebay-Salisbury, & P. S. Wells (Eds.), *The Oxford Handbook of the European Iron Age* (pp. 1–41). Oxford University Press.

Kovács, P. (2009). *Marcus Aurelius' rain miracle and the Marcomannic wars. Mnemosyne supplements 308*. Brill.

Kovács, P. (2023). *Fontes Sarmatarum in Hungaria habitantium—A magyarországi szarmatákra vonatkozó antik forrásaink*. Magyarságkutató Intézet.

Krzewińska, M., Kılınç, G. M., Juras, A., Koptekin, D., Chyleński, M., Nikitin, A. G., Shcherbakov, N., Shuteleva, I., Leonova, T., Kraeva, L., Sungatov, F. A., Sultanova, A. N., Potekhina, I., Łukasik, S., Krenz-Niedbała, M., Dalén, L., Sinika, V., Jakobsson, M., Storå, J., & Götherström, A. (2018). Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Science Advances*, *4*(10), eaat4457. https://doi.org/10.1126/sciadv.aat4457

Kulcsár, V. (1998a). *A kárpát-medencei szarmaták temetkezési szokásai*. Osváth Gedeon Múzeumi Alapítvány.

Kulcsár, V. (1998b). Újabb szempontok a hévízgyörki szarmata sírok etnikai meghatározásához. In T. Asztalos (Ed.), *Egy múzeum szolgálatában. Tanulmányok Asztalos István tiszteletére* (pp. 75–84). Osváth Gedeon Múzeumi Alapítvány.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., de Filippo, C., Prüfer, K., … Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413. https://doi.org/10.1038/nature13673

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., & Wegmann, D. (2017). ATLAS: Analysis Tools for Low-depth and Ancient Samples. *bioRxiv*. https://doi.org/10.1101/105346

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., … Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, *551*(7680), 368–372. https://doi.org/10.1038/nature24476

Maár, K., Varga, G. I. B., Kovács, B., Schütz, O., Maróti, Z., Kalmár, T., Nyerki, E., Nagy, I., Latinovics, D., Tihanyi, B., Marcsik, A., Pálfi, G., Bernert, Z., Gallina, Z., Varga, S., Költő, L., Raskó, I., Török, T., & Neparáczki, E. (2021). Maternal Lineages from 10–11th Century Commoner Cemeteries of the Carpathian Basin. *Genes*, *12*(3). https://doi.org/10.3390/genes12030460

Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., & Reich, D. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Scientific Data*, *11*(1), 182. https://doi.org/10.1038/s41597-024-03031-7

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., … Török, T. (2022). The genetic origin of Huns, Avars, and conquering Hungarians. *Current Biology*, *32*(13), 2858-2870.e7. https://doi.org/10.1016/j.cub.2022.04.093

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Masek, Z. (2016). Dák divat Pannonia határán? Az alföldi jazigok romanizációjának kezdeteihez. [Dacian fashion on the fringes of Pannonia? How the romanisation of the Jazygians of the Hungarian Plain began.]. In L. Kovács & L. Révész (Eds.), *Népek és kultúrák a Kárpát-medencében. Tanulmányok Mesterházy Károly tiszteletére* (pp. 147–178). Magyar Nemzeti Múzeum – Déri Múzeum – MTA BTK Régészeti Intézet – Szegedi Tudományegyetem.

Masek, Z. (2021). A római kori és kora középkori őséghajlati és környezettörténeti kutatások régészeti vonatkozásai. In E. Benkő & C. Zatykó (Eds.), *A Kárpát-medence környezettörténete a középkorban és a kora újkorban* (pp. 111–146). Archaeolingua Alapítvány.

Massicotte, P., & South, A. (2024). *rnaturalearth: World Map Data from Natural Earth*. https://docs.ropensci.org/rnaturalearth/

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., Fernandes, D., Ferry, M., Gamarra, B., Fortes, G. G., Haak, W., Harney, E., Jones, E., Keating, D., Krause-Kyora, B., … Reich, D. (2018). The genomic history of southeastern Europe. *Nature*, *555*(7695), 197–203. https://doi.org/10.1038/nature25778

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Melyukova, A. I. (1990). The Scythians and Sarmatians. In D. Sinor (Ed.), *The Cambridge History of Early Inner Asia* (pp. 97–117). Cambridge University Press.

Mesterházy, K. (1986). Frühsarmatenzeitlicher Grabfund aus Veresegyház. [Kora szarmata kori sírlelet Veresegyházról.]. *Folia Archaeologica*, *37*, 137–161.

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *6*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Molnár, M., Janovics, R., Major, I., Orsovszki, J., Gönczi, R., Veres, M., Leonard, A. G., Castle, S. M., Lange, T. E., Wacker, L., & al, et. (2013). Status Report of the New AMS 14C Sample Preparation Lab of the Hertelendi Laboratory of Environmental Studies (Debrecen, Hungary). *Radiocarbon*, *55*(2), 665–676. https://doi.org/10.1017/S0033822200057829

Mordvintseva, V. (2013). The Sarmatians: The Creation of Archaeological Evidence. *Oxford Journal of Achaeology*, *32*(2), 203–219. https://doi.org/10.1111/ojoa.12010

Nagy, M. (2018). *A Budapest, XVII. Rákoscsaba, Péceli úti császárkori barbár temető (Kr. U. 2-4. Század) I-II.*

Narasimhan, V. M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., Kim, A. M., Olivieri, L. M., Coppa, A., Vidale, M., Mallory, J., Moiseyev, V., Kitov, E., Monge, J., Adamski, N., … Reich, D. (2019). The formation of human populations in South and Central Asia. *Science*, *365*(6457), eaat7487. https://doi.org/10.1126/science.aat7487

Neparáczki, E., Maróti, Z., Kalmár, T., Maár, K., Nagy, I., Latinovics, D., Kustár, Á., Pálfi, G., Molnár, E., Marcsik, A., Balogh, C., Lőrinczy, G., Gál, S. S., Tomka, P., Kovacsóczy, B., Kovács, L., Raskó, I., & Török, T. (2019). Y-chromosome haplogroups from Hun, Avar and conquering Hungarian period nomadic people of the Carpathian Basin. *Scientific Reports*, *9*(1), 16569. https://doi.org/10.1038/s41598-019-53105-5

Nyerki, E., Kalmár, T., Schütz, O., Lima, R. M., Neparáczki, E., Török, T., & Maróti, Z. (2023). correctKin: An optimized method to infer relatedness up to the 4th degree from low-coverage ancient human genomes. *Genome Biology*, *24*(1), 38. https://doi.org/10.1186/s13059-023-02882-4

O'Sullivan, N., Posth, C., Coia, V., Schuenemann, V. J., Price, T. D., Wahl, J., Pinhasi, R., Zink, A., Krause, J., & Maixner, F. (2018). Ancient genome-wide analyses infer kinship structure in an Early Medieval Alemannic graveyard. *Science Advances*, *4*(9), eaao1262. https://doi.org/10.1126/sciadv.aao1262

Párducz, M. (1941). *A szarmatakor emlékei Magyarországon I./Denkmäler der Sarmatenzeit Ungarns I*. Magyar Nemzeti Múzeum.

Párducz, M. (1944). *A szarmatakor emlékei Magyarországon II./Denkmäler der Sarmatenzeit Ungarns II*. Magyar Nemzeti Múzeum.

Párducz, M. (1950). *A szarmatakor emlékei Magyarországon III./Denkmäler der Sarmatenzeit Ungarns III*. Magyar Nemzeti Múzeum.

Párducz, M. (1959). Archäologische Beitrage zur Geschichte der Hunnenzeit in Ungarn. *Acta Archaeologica Academiae Scientiarum Hungaricae*, *11*, 309–398.

Párducz, M. (1963). *Die ethnischen Probleme der Hunnenzeit in Ungarn* (Studia Arc). Akadémiai Kiadó.

Patterson, N., Isakov, M., Booth, T., Büster, L., Fischer, C.-E., Olalde, I., Ringbauer, H., Akbari, A., Cheronet, O., Bleasdale, M., Adamski, N., Altena, E., Bernardos, R., Brace, S., Broomandkhoshbacht, N., Callan, K., Candilio, F., Culleton, B., Curtis, E., … Reich, D. (2022). Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature*, *601*(7894), 588–594. https://doi.org/10.1038/s41586-021-04287-4

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*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genetics*, *2*(12), 1–20. https://doi.org/10.1371/journal.pgen.0020190

Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Pető, Á., Kenéz, Á., & Tóth, Z. (2017). Régészeti növénytani adatok a szarmaták mezőgazdaság- és gazdaság történeti kutatásához Hatvan–Baj-puszta és Apc–Farkas-major lelőhelyek

alapján/Archaeobotanical data on the economy of the sarmatians: The case study of Hatvan–Baj-puszta and Apc–Farkas-major (Heves county, Hungary). *Archeometriai Műhely*, *14*(2), 117–128.

Philippsen, B. (2013). The freshwater reservoir effect in radiocarbon dating. *Heritage Science*, *1*(1), 1–19. https://doi.org/10.1186/2050-7445-1-24

Picard toolkit. (2019). In *Broad Institute, GitHub repository*. Broad Institute. https://broadinstitute.github.io/picard/

Pinar, J. G., & Jiřík, J. (2019). Late Przeworsk and post-Przeworsk, Elbian and Danubian. Vandals, Suebi and the dissemination of Central European elements of material culture in the Western Provinces. In K. Kot-Legieć, A. Michałowskiego, M. Ol\c edzkiego, & M. Piotrowskiej (Eds.), *Przeworsk culture Transformation processes and external contacts* (pp. 405–489). Wydane przez Wydawnictwo Uniwersytetu Łódzkiego.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford Jr, T. W., Orlando, L., Metspalu, E., Karmin, M., Tambets, K., Rootsi, S., Mägi, R., Campos, P. F., Balanovska, E., Balanovsky, O., Khusnutdinova, E., Litvinov, S., … Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91. https://doi.org/10.1038/nature12736

Ralf, A., Montiel González, D., Zhong, K., & Kayser, M. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Molecular Biology and Evolution*, *35*(5), 1291–1294. https://doi.org/10.1093/molbev/msy032

Ramsey, C. B., Schulting, R., Goriunova, O. I., Bazaliiskii, V. I., & Weber, A. W. (2014). Analyzing Radiocarbon Reservoir Offsets Through Stable Nitrogen Isotopes and Bayesian Modeling: A Case Study Using Paired Human and Faunal Remains from the Cis-Baikal Region, Siberia. *Radiocarbon*, *56*(2), 789–799. Cambridge Core. https://doi.org/10.2458/56.17160

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., Kivisild, T., Zhai, W., Eriksson, A., Manica, A., Orlando, L., Vega, F. M. D. L., Tridico, S., Metspalu, E., Nielsen, K., … Willerslev, E. (2011). An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science*, *334*(6052), 94–98. https://doi.org/10.1126/science.1211177

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., & al, et. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon*, *62*(4), 725–757. https://doi.org/10.1017/RDC.2020.41

Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology*, *16*(1), 224. https://doi.org/10.1186/s13059-015-0776-0

Ringbauer, H., Huang, Y., Akbari, A., Mallick, S., Olalde, I., Patterson, N., & Reich, D. (2024). Accurate detection of identity-by-descent segments in human ancient DNA. *Nature Genetics*, *56*(1), 143–151. https://doi.org/10.1038/s41588-023-01582-w

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil – DNA – glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660). https://doi.org/10.1098/rstb.2013.0624

Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J., & Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics*, *53*(1), 120–126. https://doi.org/10.1038/s41588-020-00756-0

Schütz, O., Maróti, Z., Tihanyi, B., Kiss, A. P., Nyerki, E., Gînguță, A., Kiss, P., Varga, G. I. B., Kovács, B., Maár, K., Kovacsóczy, B. Ny., Lukács, N., Major, I., Marcsik, A., Patyi, E., Szigeti, A., Tóth, Z., Walter, D., Wilhelm, G., … Török, T. (2025). Unveiling the origins and genetic makeup of the "forgotten people": A study of the Sarmatian-period population in the Carpathian Basin. *Cell*, S0092867425005598. https://doi.org/10.1016/j.cell.2025.05.009

Soós, E. (2019a). A császárkori germán Przeworsk-kultúra kutatásának legújabb eredményei Magyarországon. *Archaeologiai Értesítő*, *144*, 67–95.

Soós, E. (2019b). Békés együttélés vagy erőszakos hódítás? Adatok a kontinuitás kérdéséhez a hun korban az északkelet-kárpát-medencei települések alapján. *Pontes*, *2*, 123–158.

Sousa da Mota, B., Rubinacci, S., Cruz Dávalos, D. I., G. Amorim, C. E., Sikora, M., Johannsen, N. N., Szmyt, M. H., Włodarczak, P., Szczepanek, A., Przybyła, M. M., Schroeder, H., Allentoft, M. E., Willerslev, E., Malaspinas, A.-S., & Delaneau, O. (2023). Imputation of ancient human genomes. *Nature Communications*, *14*(1), 3660. https://doi.org/10.1038/s41467-023-39202-0

Tari, E. (1994). Korai szarmata sír Újszilváson. [Early Sarmatian grave from Újszilvás.]. In G. Lőrinczy (Ed.), *A kőkortól a középkorig. Tanulmányok Trogmayer Ottó 60. Születésnapjára/ Von der Steinzeit bis zum Mittelalter.* (pp. 259–261). Csongrád Megyei Múzeumok Igazgatósága.

Tejral, J. (2011). *Einheimische und Fremde. Das norddanubische Gebiet zur Zeit der Völkerwanderung*. Archäologisches Institut der Akademie der Wissenschaften der Tschechischen Republik Brno.

Tokhtasyev, S. R. (2005). Sauromatae – Sarmatae – Syrmatae. *Khersonesskiy sbornik*, *14*, 291–306.

Tomka, P. (2001). The Grave of Árpás from the 5 century. *Arabona*, *39*.

Unterländer, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanová, Z., Groß, M., Sell, C., Blöcher, J., Kirsanow, K., Rohland, N., Rieger, B., Kaiser, E., Schier, W., Pozdniakov, D., Khokhlov, A., Georges, M., Wilde, S., Powell, A., Heyer, E., … Burger, J. (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nature Communications*, *8*(1), 14615. https://doi.org/10.1038/ncomms14615

Vaday, A. H. (1988). *Die sarmatischen Denkmäler des Komitats Szolnok. Ein Beitrag zur Archäologie und Geschichte des sarmatischen Barbaricums.* Archäologisches Institut der UAW.

Vaday, A. H. (2001). Military system of the Sarmatians. In: International Connection of the Barbarians of the Carpatian Basin in the 1st–5th centuries A. D. In E. Istvánovits & V. Kulcsár (Eds.), *International connections of the Barbarians of the Carpathian Basin in the 1st-5th centuries A. D. Proceedings of the International Conference held in 1999 in Aszód and Nyíregyháza.* (pp. 171–194). Jósa András Múzeum.

Varga, G. I. B., Kristóf, L. A., Maár, K., Kis, L., Schütz, O., Váradi, O., Kovács, B., Gînguță, A., Tihanyi, B., Nagy, P. L., Maróti, Z., Nyerki, E., Török, T., & Neparáczki, E. (2023). The archaeogenomic validation of Saint Ladislaus' relic provides insights into the Árpád dynasty's genealogy. *Journal of Genetics and Genomics*, *50*(1), 58–61. https://doi.org/10.1016/j.jgg.2022.06.008

Varga, S. (2020). Ásatások a Kenyere-ér partján. Egy 3. Századi germán harcos sírja a szarmaták között. In V. Csányi (Ed.), *Korok, kultúrák, lelőhelyek* (pp. 35–50). Tornyai János Múzeum.

Veeramah, K. R., Rott, A., Groß, M., Van Dorp, L., López, S., Kirsanow, K., Sell, C., Blöcher, J., Wegmann, D., Link, V., Hofmanová, Z., Peters, J., Trautmann, B., Gairhos, A., Haberstroh, J., Päffgen, B., Hellenthal, G., Haas-Gebhard, B., Harbeck, M., & Burger, J. (2018). Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proceedings of the National Academy of Sciences*, *115*(13), 3494–3499. https://doi.org/10.1073/pnas.1719880115

Wang, T., Wang, W., Xie, G., Li, Z., Fan, X., Yang, Q., Wu, X., Cao, P., Liu, Y., Yang, R., Liu, F., Dai, Q., Feng, X., Wu, X., Qin, L., Li, F., Ping, W., Zhang, L., Zhang, M., … Fu, Q. (2021). Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. *Cell*, *184*(14), 3829-3841.e21. https://doi.org/10.1016/j.cell.2021.05.018

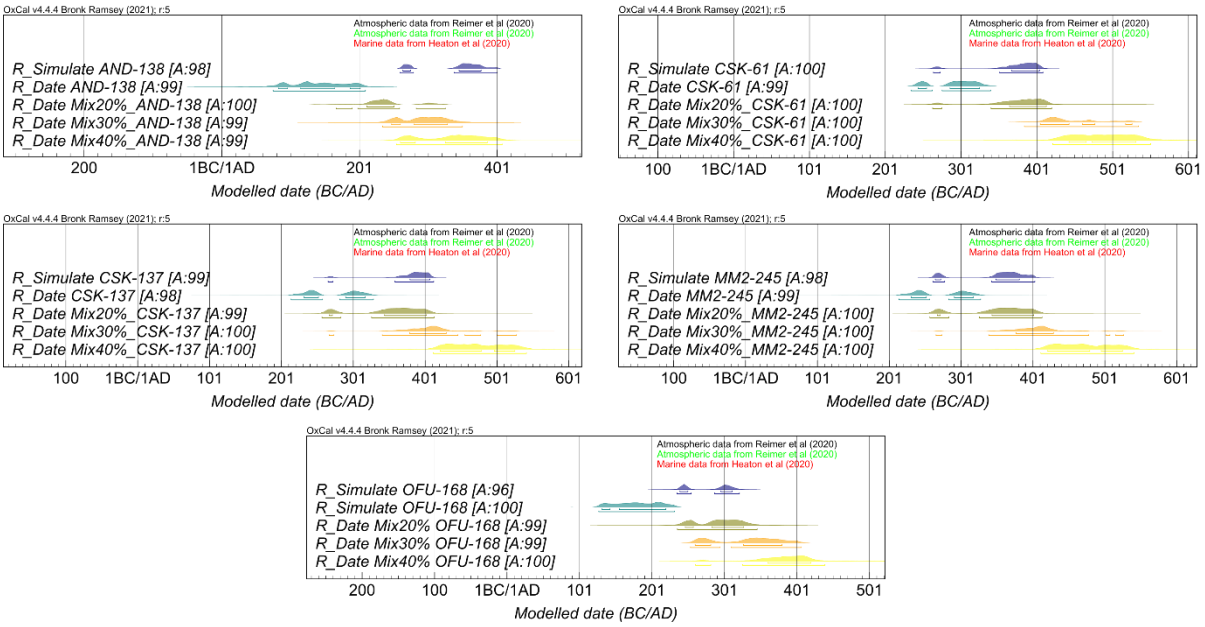
Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

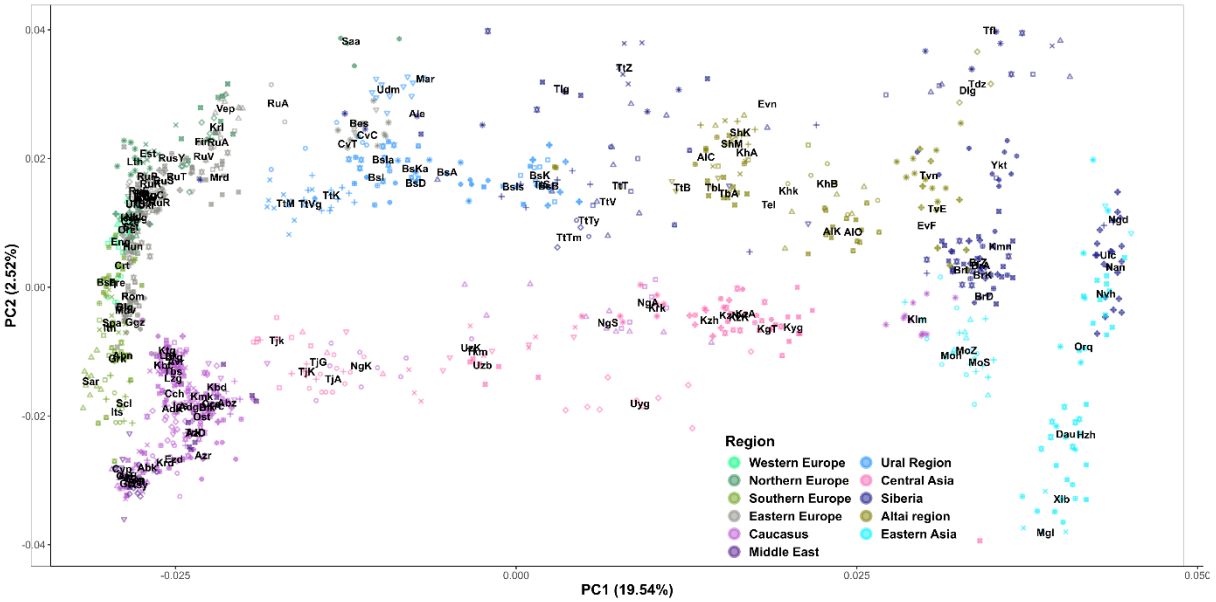
Yan, D., Li, C., Zhang, X., Wang, J., Feng, J., Dong, B., Fan, J., Wang, K., Zhang, C., Wang, H., Zhang, J., & Qin, T. (2022). A data set of global river networks and corresponding water resources zones divisions v2. *Scientific Data*, *9*(1), 770. https://doi.org/10.1038/s41597-022-01888-0

IX APPENDIX

IX.1 Figures



Extended Figure 1: Simulations for the possible impact of reservoir effect on individuals with uncertain radiocarbon dating. The expected date range ("R_Simulate") was simulated based on archaeological assessment (coloured with navy blue). The observed prior ("R_Date") distribution is indicated by teal colour. Simulated posteriors are plotted based on a proposed marine-terrestrial mixed diets of 20, 30 and 40% ("R_Date MixX", coloured with olive, orange and yellow respectively). For further details regarding the simulation see **III.1.4**.



Abk	Abkhasian	CvC	Chuvash_Cheboksary	KhA	Khakass_Askizskiy	NgS	Nogai_Stavropol	Tfl	Tofalar
Abn	Albanian	CvT	Chuvash_Tatarstan	KhB	Khakass_Beiskiy	Nvh	Nivh	TjA	Tajik_Aininsky
Abz	Abazin	Cyp	Cypriot	Khk	Khakass_Kachin	Nwg	Norwegian	TjG	Tajik_Gissar
Adg	Adygei_Shovgenovsky	Cze	Czech	Klm	Kalmyk	Orc	Orcadian	Tjk	Tajik
AdK	Adygei_Krasnodar	Ccs	Circassian	Kmk	Kumyk	Orq	Oroqen	TjK	Tajik_Kulyab
AlC	Altaian_Chelkan	Dai	Dai	Kmn	Khamnegan	Ost	Ossetian	Tkm	Turkmen
Ale	Aleut	Dau	Daur	Krc	Karachai	Rom	Romanian	Tlg	Tlingit
AlK	Altaian_KoshAgachsky	Dlg	Dolgan	Krd	Kurd	RuA	Russian_Archangelsk_Krasnoborsky	Trk	Turkish
AlO	Altaian_Ongudaisky	Dng	Dungan	Krk	Karakalpak	RuA	Russian_Archangelsk_Leshukonsky	TtB	Tatar_Irtysh_Barabinsk
ArH	Armenian_Hemsheni	Drg	Darginian	Krl	Karelian	RuB	Russian_Belgorod	TtK	Tatar_Kazan
Arm	Armenian	Eng	English	Ktg	Kaitag	RuC	Russian_Cheremisinovsky	TtM	Tatar_Mishar
Asy	Assyrian	Ent	Enets	Kyg	Kyrgyz_Kyrgyzstan	RuE	Russian_Ershichi	TtS	Tatar_Siberian_Yalutorovskiy
Avr	Avar	Est	Estonian	KzA	Kazakh_Argyn	RuK	Russian_Kaluga	TtT	Tatar_Siberian_Tobolsk
AzD	Azeri_Dagestan	EvF	Evenk_FarEast	Kzh	Kazakh	RuO	Russian_Orel	TtTm	Tatar_Tomsk
Azr	Azeri	Evn	Even	Kzh	Kazakh	RuP	Russian_Pskov	TtV	Tatar_Siberian_Vagaysky
Bes	Besermyan	EvT	Evenk_Transbaikal	KzK	Kazakh_Kypchak	RuR	Russian_Ryazan	TtVg	Tatar_Volga
Blg	Bulgarian	Ezd	Ezid	KzN	Kazakh_Naiman	RuS	Russian_Spassk	TtZ	Tatar_Siberian_Zabolotniye
Blk	Balkar	Fin	Finnish	Lah	Lahu	RusY	Russian_Yaroslavl	TtTy	Tatar_Siberian_Tyumen
Blr	Belarusian	Fre	French	Lak	Lak	RuT	Russian_Tver	Tuj	Tujia
BrA	Buryat_Aginsky	Ggz	Gagauz	Lth	Lithuanian	RuV	Russian_Vologda	Tuu	Tu
BrD	Buryat_Duldurginsky	Grg	Georgian	Lzg	Lezgin	Saa	Saami	TvE	Tuvinian_Erzin
BrI	Buryat_Irkutsk	Grk	Greek	Mar	Mari	Sar	Sardinian	Tvn	Tuvinian
BrK	Buryat_Khorinsky	GrT	Georgian_Turkey	Mdv	Moldavian	Scl	Sicilian	Udm	Udmurt
BrZ	Buryat_Zakamensky	GrZ	Georgian_Zugdidi	Mgl	Mongola	Sct	Scottish	UkL	Ukrainian_Lviv
BsA	Bashkir_Arkhangelsk	Han	Han	Mia	Miao	SeF	Selkup_Farkovo	UkN	Ukrainian_North
BsB	Bashkir_Baimaksky	Hun	Hungarian	Mns	Mansi	SeT	Selkup_Turukhansk	Ukr	Ukrainian
BsD	Bashkir_Davlekanovo	Hzh	Hezhen	Mon	Mongol	She	She	UkS	Ukrainian_Sumskaya
BsI	Bashkir_Ilishevsky	Ice	Icelandic	MoS	Mongol_Sainnoyon	ShK	Shor_Khakassia	Ulc	Ulchi
BsIa	Bashkir_Ianaul	Igs	Ingushian	MoZ	Mongol_ Zasagt	ShM	Shor_Mountain	Uyg	Uyghur
BsIs	Bashkir_Ishimbai	Ite	Itelmen	Mrd	Mordovian	Slk	Selkup	Uzb	Uzbek_Tashkent
BsK	Bashkir_Kugarchinsky	Itn	Italian_North	Nan	Nanai	Spa	Spanish	UzK	Uzbek_Khorezm
BsKa	Bashkir_Karaidelsky	Its	Italian_South	Nax	Naxi	TbA	Tubalar_Artybash	Vep	Veps
Bsq	Basque	Kbc	Kubachinian	NgA	Nogai_Astrakhan	Tbl	Tubalar_Altai	Xib	Xibo
Cch	Chechen	Kbd	Kabardinian	Ngd	Negidal	Tbs	Tabasaran	Yii	Yi
Chk	Chukchi	Ket	Ket	NgK	Nogai_Karachay_Cherkessia	Tdz	Todzin	Ykg	Yukagir
Crt	Croatian	KgT	Kyrgyz_Tajikistan	Ngn	Nganasan	Tel	Teleut	Ykt	Yakut

Extended Figure 2: Contemporary population background of 1397 modern Eurasian individuals, used to calculate the PCA axes. Ancient genomes shown on **Figures 4, 5 and Figure 8** were projected onto this background. Three-letter codes indicate the population labels positioned at the median coordinates of the corresponding individuals (for further details see Maróti et al., 2022).

IX.2 Tables

Supplementary Table 1: Archaeological and genetic data of the newly sampled individuals. The "Short Label" abbreviations are used as described in **Table 2**. Some samples exhibited signals of high exogeneous contamination or low average genome coverage that excluded them from the downstream IBD analysis. These values are marked with red in the table. For details regarding the sample classification and further sequencing data see **IV.1.3** and Schütz et al. (2025).

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
LMO-8	Lișcoteanca - Movila Olarului	M.8	ROU_IA	x	1.77	1.0	1.0	XY
RAM-7	Râmnicelu, 1969	M.7	ROU_IA	x	1.04	1.0		XX
BOT-2	Botoșani	M.2	ROU_SARM	x	2.31	1.0	1.0	XY
LMF-6	Lișcoteanca - Moș Filon	M.6	ROU_SARM		1.99	1.0		XX
LMF-7	Lișcoteanca - Moș Filon	M.7	ROU_SARM	x	1.47	1.0	0.4	XY
OLR-2	Olteniță - Renie, 1968	M.2	ROU_SARM	x	0.58	1.0	2.8	XY
OSU-1	Olteniță - Secția Ulmeni, 1960	M.1	ROU_SARM	x	1.24	1.0		XX
POG-10	Pogorăști	M.10	ROU_SARM	x	1.47	1.0	0.6	XY
PRO-37	Probota	M.37/1960	ROU_SARM	x	1.39	1.0		XX
PRO-47	Probota	M.47/1961	ROU_SARM	x	1.40	1.0		XX
RAM-13	Râmnicelu, 1969	M.13	ROU_SARM		1.31	**10.0**	0.6	XY
RLS-1	Ripiceni - La Stâncă	M.1	ROU_SARM	x	1.47	1.0		XX
TAR-118	Târgșor, 1960	M.118	ROU_SARM	x	1.21	1.0	2.6	XY
TAR-184	Târgșor, 1960	M.184	ROU_SARM		1.36	1.0		XX
TAR-196	Târgșor, 1960	M.196	ROU_SARM	x	1.24	1.0	1.0	XY
TAF-11	Târgu Frumos	T.1/M.11	ROU_SARM	x	0.91	1.0	1.4	XY
TRE-33	Trestiana	M.33/1972	ROU_SARM	x	1.01	1.0		XX
DZS-1	Dormánd - Zsidótemető	1	HUN_SARM_EP		1.45	1.0		XX
DZS-3	Dormánd - Zsidótemető	3	HUN_SARM_EP		1.32	1.0	1.0	XY
DZS-5	Dormánd - Zsidótemető	5	HUN_SARM_EP		1.39	1.0	1.0	XY

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
DZS-6	Dormánd - Zsidótemető	6	HUN_SARM_EP		1.53	1.0		XX
DZS-41	Dormánd - Zsidótemető	41	HUN_SARM_EP	x	0.90	2.0	0.6	XY
DZS-42	Dormánd - Zsidótemető	42	HUN_SARM_EP	x	1.01	1.0		XX
DZS-43	Dormánd - Zsidótemető	43	HUN_SARM_EP		1.07	1.0	1.1	XY
DZS-44	Dormánd - Zsidótemető	44	HUN_SARM_EP		1.15	1.0	0.9	XY
DZS-48	Dormánd - Zsidótemető	48	HUN_SARM_EP		2.51	1.0		XX
FKD-60	Füzesabony - Kastélydűlő	60	HUN_SARM_EP	x	1.59	1.0		XX
FKD-140	Füzesabony - Kastélydűlő	140	HUN_SARM_EP		0.98	1.0		XX
FKD-150	Füzesabony - Kastélydűlő	150	HUN_SARM_EP	x	1.34	1.0		XX
MEH-2p	Mezőcsát - Hörcsögös	pit 2	HUN_SARM_EP	x	1.50	1.0		XX
CSU2-39	Szeged - Csongrádi út	39	HUN_SARM_EP		1.31	1.0	1.0	XY
CSU2-43	Szeged - Csongrádi út	43	HUN_SARM_EP	x	1.56	1.0	1.3	XY
HVF-2	Hódmezővásárhely - Fehértó	2	HUN_SARM_EMP		1.35	2.0		XX
HVF-3	Hódmezővásárhely - Fehértó	3	HUN_SARM_EMP		1.43	1.0		XX
HVF-4	Hódmezővásárhely - Fehértó	4	HUN_SARM_EMP	x	1.14	1.0	0.9	XY
HVF-8	Hódmezővásárhely - Fehértó	8	HUN_SARM_EMP		1.43	1.0		XX
HVF-10	Hódmezővásárhely - Fehértó	10	HUN_SARM_EMP		1.19	1.0		XX
HVF-15	Hódmezővásárhely - Fehértó	15	HUN_SARM_EMP		0.87	1.0		XX
HVF-17	Hódmezővásárhely - Fehértó	17	HUN_SARM_EMP		1.04	1.0	0.9	XY
HVF-21	Hódmezővásárhely - Fehértó	21	HUN_SARM_EMP	x	2.09	1.0	0.4	XY
KAK-3	Kunadacs - Községi temető	3	HUN_SARM_EMP	x	1.14	3.0		XX
MIJ-1	Makó - Igási Járandó	1	HUN_SARM_EMP	x	1.84	1.0	1.1	XY
MIJ-3	Makó - Igási Járandó	3	HUN_SARM_EMP	x	1.30	1.0	1.3	XY
MIJ-4	Makó - Igási Járandó	4	HUN_SARM_EMP	x	1.75	**9.0**		XX
MIJ-7	Makó - Igási Járandó	7	HUN_SARM_EMP	x	1.55	1.0		XX
AND-128	Apátfalva - Nagyút dűlő M43, 43. lh	128	HUN_SARM_MLP		1.43	0.0	0.3	XY

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
AND-131	Apátfalva - Nagyút dűlő M43, 43. lh	131	HUN_SARM_MLP		3.75	1.0	0.3	XY
AND-138	Apátfalva - Nagyút dűlő M43, 43. lh	138	HUN_SARM_MLP	x	1.86	1.0	0.6	XY
AND-145	Apátfalva - Nagyút dűlő M43, 43. lh	145	HUN_SARM_MLP		1.41	1.0		XX
AND-177	Apátfalva - Nagyút dűlő M43, 43. lh	177	HUN_SARM_MLP		1.11	1.0		XX
AFM-213	Apc - Farkas-major	213. obj.	HUN_SARM_MLP		1.49	1.0		XX
AFM-273	Apc - Farkas-major	273. obj.	HUN_SARM_MLP		0.85	0.0	1.0	XY
AFM-590	Apc - Farkas-major	590. obj	HUN_SARM_MLP		1.02	1.0		XX
AFM-940	Apc - Farkas-major	940. obj.	HUN_SARM_MLP		1.22	1.0		XX
AFM-1156	Apc - Farkas-major	1156. obj	HUN_SARM_MLP	x	0.96	1.0		XX
AFM-1255	Apc - Farkas-major	1255. obj	HUN_SARM_MLP		1.31	1.0	0.8	XY
AFM-1861	Apc - Farkas-major	1861. obj.	HUN_SARM_MLP	x	1.59	2.0		XX
RAK-115	Budapest XVII. - Rákoscsaba, Péceli út	115	HUN_SARM_MLP		1.68	1.0		XX
RAK-136	Budapest XVII. - Rákoscsaba, Péceli út	136	HUN_SARM_MLP		1.60	1.0		XX
RAK-211	Budapest XVII. - Rákoscsaba, Péceli út	211	HUN_SARM_MLP		1.76	1.0	1.0	XY
RAK-251	Budapest XVII. - Rákoscsaba, Péceli út	251	HUN_SARM_MLP		1.57	1.0		XX
HGY-14	Hévízgyörk	14	HUN_SARM_MLP	x	2.01	1.0		XX
HKB-309	Hódmezővásárhely - Kenyere-ér	309	HUN_SARM_MLP	x	1.93	1.0	0.3	XY
KFH-2	Kiskőrös - Fekete halom	2	HUN_SARM_MLP		1.66	1.0	0.9	XY
KFH-3	Kiskőrös - Fekete halom	3	HUN_SARM_MLP		1.42	1.0		XX
KDS-121	Kiskundorozsma - Subasa 26/78	121	HUN_SARM_MLP		0.97	1.0		XX
KDS-138	Kiskundorozsma - Subasa 26/79	138	HUN_SARM_MLP		1.26	1.0	0.8	XY
KDS-148	Kiskundorozsma - Subasa 26/80	148	HUN_SARM_MLP	x	1.35	1.0	0.4	XY
KDS-260	Kiskundorozsma - Subasa 26/81	260	HUN_SARM_MLP		0.67	1.0	0.7	XY
MDH-28	Madaras - Halmok	28	HUN_SARM_MLP		1.43	1.0	0.6	XY
MDH-48	Madaras - Halmok	48	HUN_SARM_MLP		**0.36**	2.0	4.3	XY
MDH-82	Madaras - Halmok	82	HUN_SARM_MLP	x	1.58	1.0	0.4	XY

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
MDH-84	Madaras - Halmok	84	HUN_SARM_MLP		0.62	1.0		XX
MDH-162	Madaras - Halmok	162	HUN_SARM_MLP		2.17	1.0		XX
MDH-178	Madaras - Halmok	178	HUN_SARM_MLP		1.34	0.0	0.8	XY
MDH-188	Madaras - Halmok	188	HUN_SARM_MLP		1.39	1.0		XX
MDH-209	Madaras - Halmok	209	HUN_SARM_MLP		2.01	1.0		XX
MDH-221	Madaras - Halmok	221	HUN_SARM_MLP		1.23	1.0		XX
MDH-249	Madaras - Halmok	249	HUN_SARM_MLP	x	1.25	1.0	0.6	XY
MDH-263	Madaras - Halmok	263	HUN_SARM_MLP		1.40	1.0	0.9	XY
MDH-265	Madaras - Halmok	265	HUN_SARM_MLP	x	1.65	1.0	0.9	XY
MDH-291	Madaras - Halmok	291	HUN_SARM_MLP		1.43	1.0		XX
MDH-342	Madaras - Halmok	342	HUN_SARM_MLP		3.24	1.0		XX
MDH-344	Madaras - Halmok	344	HUN_SARM_MLP	x	1.17	1.0		XX
MDH-357	Madaras - Halmok	357	HUN_SARM_MLP	x	1.98	1.0	0.9	XY
MDH-405	Madaras - Halmok	405	HUN_SARM_MLP		**0.24**	2.0	8.8	XY
MDH-410	Madaras - Halmok	410	HUN_SARM_MLP		1.42	1.0		XX
MDH-444	Madaras - Halmok	444	HUN_SARM_MLP		1.52	1.0	0.8	XY
MDH-462	Madaras - Halmok	462	HUN_SARM_MLP		0.54	1.0	5.6	XY
MDH-483	Madaras - Halmok	483	HUN_SARM_MLP		1.54	1.0		XX
MDH-500	Madaras - Halmok	500	HUN_SARM_MLP		**0.37**	3.0		XX
MDH-514	Madaras - Halmok	514	HUN_SARM_MLP		1.64	0.0	0.9	XY
MDH-592	Madaras - Halmok	592	HUN_SARM_MLP	x	1.08	1.0	1.3	XY
MDH-630	Madaras - Halmok	630	HUN_SARM_MLP		1.63	1.0	0.6	XY
MDH-656	Madaras - Halmok	656	HUN_SARM_MLP		1.79	1.0		XX
MDH-661	Madaras - Halmok	661	HUN_SARM_MLP		1.24	2.0	0.6	XY
PLG-1	Püspökladány - Görepart	1	HUN_SARM_MLP		1.66	1.0	1.0	XY
PLG-4	Püspökladány - Görepart	4	HUN_SARM_MLP	x	0.72	1.0	5.1	XY

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
PLG-5	Püspökladány - Görepart	5	HUN_SARM_MLP		1.37	1.0	0.9	XY
ZZ-1	Zákányszék - Zákánydűlő NY/69	1	HUN_SARM_MLP	x	1.62	1.0		XX
CSO-502	Csanádpalota - Országhatár	50	HUN_SARM_LP	x	1.04	1.0	0.9	XY
CSO-507	Csanádpalota - Országhatár	40	HUN_SARM_LP		1.08	1.0	0.5	XY
CSO-526	Csanádpalota - Országhatár	42	HUN_SARM_LP		1.89	1.0		XX
MM2-215	Makó - Mikócsa dűlő M43, 31. lh.	215	HUN_SARM_LP		1.71	1.0		XX
MM2-245	Makó - Mikócsa dűlő M43, 31. lh.	245	HUN_SARM_LP	x	1.27	2.0		XX
OFU-15	Óföldeák - Ürmös II M43 10. lh.	15	HUN_SARM_LP		1.49	1.0	0.5	XY
OFU-26	Óföldeák - Ürmös II M43 10. lh.	26	HUN_SARM_LP		1.61	2.0	0.8	XY
OFU-27	Óföldeák - Ürmös II M43 10. lh.	27	HUN_SARM_LP		1.43	0.0		XX
OFU-93	Óföldeák - Ürmös II M43 10. lh.	93	HUN_SARM_LP		1.40	2.0		XX
OFU-138	Óföldeák - Ürmös II M43 10. lh.	138	HUN_SARM_LP	x	1.45	1.0	0.6	XY
OFU-168	Óföldeák - Ürmös II M43 10. lh.	168	HUN_SARM_LP	x	1.09	1.0	0.7	XY
SPT-23	Szihalom - Pamlényi tábla	23	HUN_SARM_LP	x	1.08	1.0		XX
TMA-2	Tápé - Malajdok A	2	HUN_SARM_LP		0.69	1.0	1.1	XY
TMA-8	Tápé - Malajdok A	8	HUN_SARM_LP		1.17	0.0	0.9	XY
TMA-9	Tápé - Malajdok A	9	HUN_SARM_LP	x	1.29	1.0	0.7	XY
TMA-10	Tápé - Malajdok A	10	HUN_SARM_LP		1.05	1.0	1.0	XY
TMA-36	Tápé - Malajdok A	36	HUN_SARM_LP	x	1.10	1.0		XX
TD-SZ	Tiszadob - Sziget	stray find	HUN_SARM_LP	x	0.96	1.0	0.4	XY
TIV-6	Tiszavalk	6	HUN_SARM_LP		1.48	1.0	0.8	XY
TIV-9	Tiszavalk	9	HUN_SARM_LP		1.65	1.0	0.9	XY
TIV-13	Tiszavalk	13	HUN_SARM_LP	x	1.75	1.0		XX
TIV-17	Tiszavalk	17	HUN_SARM_LP		1.62	2.0	1.1	XY
KM-747	Kecskemét - Mercedes RL 15	SNR747	HUN_SARM_UP		1.20	1.0		XX
KM-1005	Kecskemét - Mercedes RL 15	SNR1005	HUN_SARM_UP	x	1.21	3.0		XX

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
KM-3679	Kecskemét - Mindszenti-dűlő I	SNR3679/B	HUN_SARM_UP	x	1.27	1.0	0.8	XY
KM-156	Kecskemét - Mindszenti-dűlő II	SNR156	HUN_SARM_UP		0.98	1.0		XX
KM-199	Kecskemét - Mindszenti-dűlő II	SNR199	HUN_SARM_UP		1.86	1.0		XX
KSZN-10321	Kunszentmiklós - Nyakvágócsárda	-	HUN_SARM_UP	x	1.89	1.0	0.8	XY
NKL-7	Nagykálló - Kis Ludas-tó dűlő	7	HUN_SARM_UP	x	1.23	1.0	0.8	XY
NKL-112	Nagykálló - Kis Ludas-tó dűlő	112	HUN_SARM_UP		0.95	1.0		XX
NKL-135	Nagykálló - Kis Ludas-tó dűlő	135	HUN_SARM_UP	x	1.45	1.0		XX
NKL-157	Nagykálló - Kis Ludas-tó dűlő	157	HUN_SARM_UP	x	1.38	1.0		XX
NKT-20	Nemesnádudvar - Külbogyiszló	20	HUN_SARM_UP	x	1.09	1.0		XX
NKT-32	Nemesnádudvar - Külbogyiszló	32	HUN_SARM_UP		1.74	1.0	0.5	XY
RBR-118	Rákóczifalva - Bivaly-tó, Rokkant-föld	118	HUN_SARM_UP	x	1.01	1.0	1.1	XY
CSK-9	Csongrád - Kenderföldek	9	HUN_SARM_HUN	x	1.40	1.0	0.6	XY
CSK-18	Csongrád - Kenderföldek	18	HUN_SARM_HUN		1.90	1.0		XX
CSK-25	Csongrád - Kenderföldek	25	HUN_SARM_HUN		1.78	1.0	0.7	XY
CSK-61	Csongrád - Kenderföldek	61	HUN_SARM_HUN	x	1.79	1.0		XX
CSK-101	Csongrád - Kenderföldek	101	HUN_SARM_HUN		2.03	**6.0**	0.7	XY
CSK-126	Csongrád - Kenderföldek	126	HUN_SARM_HUN		2.48	1.0		XX
CSK-127	Csongrád - Kenderföldek	127	HUN_SARM_HUN		2.67	1.0		XX
CSK-135	Csongrád - Kenderföldek	135	HUN_SARM_HUN		1.73	1.0	0.9	XY
CSK-137	Csongrád - Kenderföldek	137	HUN_SARM_HUN	x	1.40	1.0	0.9	XY
SZB-1	Szabadszállás - Boczka tanya	1	HUN_SARM_HUN	x	1.08	1.0	1.1	XY
KAP-2	Kapolcs	2	HUN_HUN		1.99	2.0		XX
MEM-2	Mezőkövesd - Mocsolyás	2	HUN_HUN	x	1.04	1.0	0.9	XY
NKL-899	Nagykálló - Kis Ludas-tó dűlő	899	HUN_HUN	x	0.87	1.0		XX
NKL-1038	Nagykálló - Kis Ludas-tó dűlő	1038	HUN_HUN		2.32	1.0		XX
OFU-190	Óföldeák - Ürmös M43 9-10. lh.	190	HUN_HUN	x	2.00	1.0		XX

Master ID	Cemetery	Grave no.	Short Label	Radiocarbon dating	Avg. coverage (fold)	Estimated MT contamination (%, Schmutzi)	Estimated X contamination (%, ANGSD)	Sex determination (Skoglund)
OFU-191	Óföldeák - Ürmös M43 9-10. lh.	191	HUN_HUN		1.49	1.0		XX
OFU-422	Óföldeák - Ürmös M43 9-10. lh.	422	HUN_HUN	x	1.57	1.0	0.7	XY
PMD-564	Páty - Malom-dűlő, 9. lh.	564	HUN_HUN		0.86	1.0		XX
PM-5	Pécs - Málom	5	HUN_HUN		1.49	1.0		XX
SPF-1	Solt - Polya-fok	1. obj (SNR15)	HUN_HUN	x	0.97	0.0		XX
VIG-94-1	Visegrád - Gizellamajor	94./1.	HUN_HUN	x	2.10	**18.0**		XX

Supplementary Table 2: Archaeological and genetic data of the previously published individuals analysed in this study. The "Short Label" abbreviations are used as described in **Table 2**. The average genome coverage for the 1240K SNP capture samples was calculated based on the 1240K marker set. The reference format was lifted from the Allen Ancient DNA Resource (AADR, Mallick et al., 2024). For details regarding the sample classification and further sequencing data see **IV.1.3** and Schütz et al. (2025).

Master ID	Cemetery	Grave no.	Short Label	Sequence information	Avg. coverage (fold)	Sex determination (Skoglund)	Reference AADR
I20802	Derecske - Karakas dűlő	54/54	HUN_SARM_MLP	1240K SNP	4.76	XY	GnecchiRusconeCell2022
A181013	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	0.24	XX	GnecchiRusconeCell2022
A181014	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	3.42	XY	GnecchiRusconeCell2022
A181015	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	3.28	XY	GnecchiRusconeCell2022
A181016	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	0.30	XY	GnecchiRusconeCell2022
A181017	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	0.43	XY	GnecchiRusconeCell2022
A181018	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	3.79	XX	GnecchiRusconeCell2022
A181019	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	1.45	XY	GnecchiRusconeCell2022
A181020	Kecskemét - Mindszenti-dűlő	-	HUN_SARM_LP	1240K SNP	0.69	XX	GnecchiRusconeCell2022
A181021	Hajdúnánás - Fürj-halom-dűlő 2. site	3/3	HUN_SARM_LP	1240K SNP	3.63	XX	GnecchiRusconeCell2022
A181022	Hajdúnánás - Fürj-halom-dűlő 2. site	9/12	HUN_SARM_LP	1240K SNP	2.38	XX	GnecchiRusconeCell2022
A181023	Hajdúnánás - Fürj-halom-dűlő 2. site	18/35	HUN_SARM_LP	1240K SNP	3.26	XY	GnecchiRusconeCell2022
A181024	Hajdúnánás - Fürj-halom-dűlő 2. site	23/41	HUN_SARM_LP	1240K SNP	3.93	XX	GnecchiRusconeCell2022
A181025	Hajdúnánás - Fürj-halom-dűlő 2. site	27/46	HUN_SARM_LP	1240K SNP	3.66	XY	GnecchiRusconeCell2022
A181026	Hajdúnánás - Fürj-halom-dűlő 2. site	36/64	HUN_SARM_LP	1240K SNP	3.69	XX	GnecchiRusconeCell2022
A181027	Hajdúnánás - Fürj-halom-dűlő 2. site	43/74	HUN_SARM_LP	1240K SNP	3.46	XX	GnecchiRusconeCell2022
A181028	Hajdúnánás - Fürj-halom-dűlő 2. site	44/75	HUN_SARM_LP	1240K SNP	2.93	XY	GnecchiRusconeCell2022
CSB-3	Csongrád - Berzsenyi utca	3	HUN_HUN	Shotgun	1.19	XY	MarotiTorokCurrBio2022
VZ-12673	Budepest - Vezér utca	solitary	HUN_HUN	Shotgun	3.58	XY	MarotiTorokCurrBio2022
MSG-1	Marosszentgyörgy - Kerekdomb	1	HUN_HUN	Shotgun	1.51	XY	MarotiTorokCurrBio2022

Master ID	Cemetery	Grave no.	Short Label	Sequence information	Avg. coverage (fold)	Sex determination (Skoglund)	Reference AADR
KMT-2785	Kecskemét - Mindszenti-dűlő-RL	solitary	HUN_HUN	Shotgun	2.12	XY	MarotiTorokCurrBio2022
ASZK-1	Árpás - Dombiföld, Szérűskert	1	HUN_HUN	Shotgun	2.01	XY	MarotiTorokCurrBio2022
SEI-1	Sándorfalva - Eperjes, Ivótavak	1	HUN_HUN	Shotgun	1.19	XY	MarotiTorokCurrBio2022
SEI-5	Sándorfalva - Eperjes, Ivótavak	5	HUN_HUN	Shotgun	1.21	XY	MarotiTorokCurrBio2022
SEI-6	Sándorfalva - Eperjes, Ivótavak	6	HUN_HUN	Shotgun	1.45	XX	MarotiTorokCurrBio2022
SZLA-646	Szilvásvárad - Lovaspálya	646	HUN_HUN	Shotgun	1.25	XX	MarotiTorokCurrBio2022

Supplementary Table 3: Calibrated radiocarbon dates of the selected 68 individuals. Close kin connections were gathered according to the IBD analysis results (for further details see section **III.2.6** and **IV.5**).

Master ID	Archaeologocal context	Median Age (BP)	Calibrated radiocarbon dating (BCE/CE)	Close connections IBD (≥ 80 cM)
LMO-8	Western Steppe Iron Age	2727	801-753 (84.5%) calBCE	
RAM-7	Western Steppe Iron Age	2593	593-461 (38.9%), 752-683 (36.7%), 669-632 (15%) calBCE	
BOT-2	Western Steppe Sarmatian Period	1540	361-434 (93.2%) calCE	
LMF-7	Western Steppe Sarmatian Period	1865	25-125 (95.4%) calCE	
OLR-2	Western Steppe Sarmatian Period	2106	202-95 (80.1%) calBCE	
OSU-1	Western Steppe Sarmatian Period	2038	161-44 (95.4%) calBCE	
POG-10	Western Steppe Sarmatian Period	1930	43 calBCE - 77 calCE (95.4%)	
PRO-37	Western Steppe Sarmatian Period	1736	154-242 (89.6%) calCE	PRO-47
PRO-47	Western Steppe Sarmatian Period	1832	66-205 (95.4%) calCE	PRO-37
RLS-1	Western Steppe Sarmatian Period	1868	24-125 (95.4%) calCE	
TAR-118	Western Steppe Sarmatian Period	1839	62-205 (95.4%) calCE	
TAR-196	Western Steppe Sarmatian Period	1963	53 calBCE - 29 calCE (92%)	
TAF-11	Western Steppe Sarmatian Period	1794	116-211 (90%) calCE	
TRE-33	Western Steppe Sarmatian Period	1467	430-542 (95.4%) calCE	
DZS-41	Carpathian Basin Early Sarmatian Period	1858	22-130 (94.2%) calCE	
DZS-42	Carpathian Basin Early Sarmatian Period	1894	7-120 (94.2%) calCE	
FKD-60	Carpathian Basin Early Sarmatian Period	1902	5-85 (76.9%), 94-118 (12.5%) calCE	
FKD-150	Carpathian Basin Early Sarmatian Period	1788	117-215 (92.6%) calCE	
MEH-2p	Carpathian Basin Early Sarmatian Period	1726	201-249 (87%) calCE	
CSU2-43	Carpathian Basin Early Sarmatian Period	1892	9-121 (95.4%) calCE	
HVF-4	Carpathian Basin Early-Middle Sarmatian Period	1797	112-211 (87.7%) calCE	
HVF-21	Carpathian Basin Early-Middle Sarmatian Period	1775	126-225 (95.4%) calCE	
KAK-3	Carpathian Basin Early-Middle Sarmatian Period	1821	69-208 (95.4%) calCE	
MIJ-1	Carpathian Basin Early-Middle Sarmatian Period	1781	120-226 (95.4%) calCE	

Master ID	Archaeologocal context	Median Age (BP)	Calibrated radiocarbon dating (BCE/CE)	Close connections IBD (≥ 80 cM)
MIJ-3	Carpathian Basin Early-Middle Sarmatian Period	1896	7-120 (91.7%) calCE	
MIJ-4	Carpathian Basin Early-Middle Sarmatian Period	1784	110-236 (90.3%) calCE	
MIJ-7	Carpathian Basin Early-Middle Sarmatian Period	1790	72-236 (95.4%) calCE	AND-145, HVF-2
AND-138	Carpathian Basin Middle-Late Sarmatian Period	1810	75-209 (95.4%) calCE	AND-128, AND-131
AFM-1156	Carpathian Basin Middle-Late Sarmatian Period	1738	152-242 (87.6%) calCE	
AFM-1861	Carpathian Basin Middle-Late Sarmatian Period	1737	152-243 (88%) calCE	
HGY-14	Carpathian Basin Middle-Late Sarmatian Period	1777	124-220 (95.4%) calCE	
HKB-309	Carpathian Basin Middle-Late Sarmatian Period	1730	153-247 (91.3%) calCE	
KDS-148	Carpathian Basin Middle-Late Sarmatian Period	1790	115-215 (90.8%) calCE	
MDH-82	Carpathian Basin Middle-Late Sarmatian Period	1652	276-339 (75.5%), 241-262 (20%) calCE	
MDH-249	Carpathian Basin Middle-Late Sarmatian Period	1654	281-329 (66%), 234-258 (29.4%) calCE	
MDH-265	Carpathian Basin Middle-Late Sarmatian Period	1647	271-350 (76.7%), 244-266 (18.8%) calCE	
MDH-344	Carpathian Basin Middle-Late Sarmatian Period	1709	213-254 (67.6%), 289-320 (27.9%) calCE	
MDH-357	Carpathian Basin Middle-Late Sarmatian Period	1645	245-353 (95.4%) calCE	
MDH-592	Carpathian Basin Middle-Late Sarmatian Period	1745	153-238 (86.6%) calCE	
PLG-4	Carpathian Basin Middle-Late Sarmatian Period	1653	280-330 (70.4%), 237-259 (25%) calCE	PLG-5
ZZ-1	Carpathian Basin Middle-Late Sarmatian Period	1769	127-233 (95.4%) calCE	
CSO-502	Carpathian Basin Late Sarmatian Period	1652	276-343 (72.2%), 236-263 (23.2%) calCE	
MM2-245	Carpathian Basin Late Sarmatian Period	1671	283-327 (48%), 215-257 (47.4%) calCE	MM2-215
OFU-138	Carpathian Basin Late Sarmatian Period	1655	278-335 (65.2%), 230-260 (30.3%) calCE	
OFU-168	Carpathian Basin Late Sarmatian Period	1771	127-232 (95.4%) calCE	TMH-509, TMH-798
SPT-23	Carpathian Basin Late Sarmatian Period	1603	325-406 (68.7%), 255-286 (26.8%) calCE	
TMA-9	Carpathian Basin Late Sarmatian Period	1588	330-410 (76.4%), 258-281 (19.1%) calCE	
TMA-36	Carpathian Basin Late Sarmatian Period	1583	335-412 (78.7%), 259-280 (16.8%) calCE	
TD-SZ	Carpathian Basin Late Sarmatian Period	1632	246-380 (95.4%) calCE	
TIV-13	Carpathian Basin Late Sarmatian Period	1613	316-405 (63.1%), 251-294 (32.3%) calCE	TIV-6
KM-1005	Carpathian Basin Sarmatian Unkown Period	1647	242-361 (95.4%) calCE	

Master ID	Archaeologocal context	Median Age (BP)	Calibrated radiocarbon dating (BCE/CE)	Close connections IBD (≥ 80 cM)
KM-3679	Carpathian Basin Sarmatian Unkown Period	1575	339-415 (82.3%), 261-279 (13.1%) calCE	KM-747
KSZN-10321	Carpathian Basin Sarmatian Unkown Period	1837	61-206 (94.1%) calCE	
NKL-7	Carpathian Basin Sarmatian Unkown Period	1788	116-216 (92.2%) calCE	
NKL-135	Carpathian Basin Sarmatian Unkown Period	1854	25-133 (87.9%) calCE	
NKL-157	Carpathian Basin Sarmatian Unkown Period	1716	203-256 (76.6%), 285-325 (18.4%) calCE	
NKT-20	Carpathian Basin Sarmatian Unkown Period	1803	78-210 (95.4%) calCE	
RBR-118	Carpathian Basin Sarmatian Unkown Period	1634	247-377 (95.4%) calCE	
CSK-9	Carpathian Basin Late Sarmatian-Hun Period	1548	351-426 (91.7%) calCE	
CSK-61	Carpathian Basin Late Sarmatian-Hun Period	1653	276-340 (70.8%), 235-262 (24.6%) calCE	
CSK-137	Carpathian Basin Late Sarmatian-Hun Period	1666	283-328 (49.5%), 215-257 (46%) calCE	
SZB-1	Carpathian Basin Late Sarmatian-Hun Period	1645	242-363 (95.4%) calCE	
MEM-2	Carpathian Basin Hun Period	1474	418-484 (52.7%), 489-538 (42.7%) calCE	
NKL-899	Carpathian Basin Hun Period	1576	338-415 (81.9%), 260-279 (13.6%) calCE	
OFU-190	Carpathian Basin Hun Period	1471	420-538 (95.4%) calCE	
OFU-422	Carpathian Basin Hun Period	1573	341-415 (83%), 261-278 (12.4%) calCE	TMH-199, SZOD1-829
SPF-1	Carpathian Basin Hun Period	1474	418-538 (95.4%) calCE	
VIG-94-1	Carpathian Basin Hun Period	1469	424-539 (95.4%) calCE	

Supplementary Table 4: Kinship groups of the studied individuals identified by correctKin (Nyerki et al., 2023). The missing Y-chromosome haplogroup of female individuals is signified by a "XX" sign. Further details regarding the kinship analysis and complementary statistical data can be found in section **III.2.6** and Schütz et al. (2025).

Kin1 Master ID	Kin1 archaeological context	Kin2 Master ID	Kin2 archaeological context	Mt Hg Kin1 - Kin2	Y Hg Ki1 - Kin2	Corrected kin coefficient	Estimated relatedness
SEI-1	Carpathian Basin Hun Period	SEI-5	Carpathian Basin Hun Period	T1a1 - T1a1	R1a1a1b1a2b3a1a~ - R1a1a1b1a2b3a1~	0.278	1st degree
A181014	Carpathian Basin Late Sarmatian Period	A181019	Carpathian Basin Late Sarmatian Period	K1a4a1 - K1a4a1	I2a1b1a2b1~ - I2a1b1a2b1a~	0.290	1st degree
KM-3679	Carpathian Basin Sarmatian Unknown Period	KM-747	Carpathian Basin Sarmatian Unknown Period	H7b1 - H7b1	I1a2a1a1a1b - XX	0.235	1st degree
PLG-4	Carpathian Basin Middle-Late Sarmatian Period	PLG-5	Carpathian Basin Middle-Late Sarmatian Period	H6a1b - H6a1b	R1b1a1b1a1a2a1a - I2a1b1a2a1b~	0.108	2nd degree
A181015	Carpathian Basin Late Sarmatian Period	A181016	Carpathian Basin Late Sarmatian Period	U2e1h - H1cf	I1a2a1a1a1b - I1a2a1a1a	0.140	2nd degree
A181015	Carpathian Basin Late Sarmatian Period	A181017	Carpathian Basin Late Sarmatian Period	U2e1h - H7	I1a2a1a1a1b - I1a	0.108	2nd degree
A181016	Carpathian Basin Late Sarmatian Period	A181017	Carpathian Basin Late Sarmatian Period	H1cf - H7	I1a2a1a1a - I1a	0.081	3rd degree
MDH-656	Carpathian Basin Middle-Late Sarmatian Period	MDH-661	Carpathian Basin Middle-Late Sarmatian Period	T2+16189 - U5b1b1+@16192	XX - R1a1a1b2a2b2b2~	0.246	1st degree
MDH-661	Carpathian Basin Middle-Late Sarmatian Period	MDH-48	Carpathian Basin Middle-Late Sarmatian Period	U5b1b1+@16192 - H2a2b1	R1a1a1b2a2b2b2~ - R1a1a1b2a2b2~	0.054	3rd degree
MDH-656	Carpathian Basin Middle-Late Sarmatian Period	MDH-48	Carpathian Basin Middle-Late Sarmatian Period	T2+16189 - H2a2b1	XX - R1a1a1b2a2b2~	0.020	uncertain
AND-128	Carpathian Basin Middle-Late Sarmatian Period	AND-138	Carpathian Basin Middle-Late Sarmatian Period	N1a1a1a2 - N1a1a1a2	R1a1a1b2a2b2b2~ - R1a1a1b2a2b2b2~	0.246	1st degree
AND-128	Carpathian Basin Middle-Late Sarmatian Period	AND-131	Carpathian Basin Middle-Late Sarmatian Period	N1a1a1a2 - H11a2	R1a1a1b2a2b2b2~ - R1a1a1b2a2b2b2~	0.025	4th degree
AND-131	Carpathian Basin Middle-Late Sarmatian Period	AND-138	Carpathian Basin Middle-Late Sarmatian Period	H11a2 - N1a1a1a2	R1a1a1b2a2b2b2~ - R1a1a1b2a2b2b2~	0.026	4th degree
DZS-3	Carpathian Basin Early Sarmatian Period	DZS-5	Carpathian Basin Early Sarmatian Period	X2b6a - U5a2+16294	E1b1b1a1b1a - E1b1b1a1b1a	0.092	3rd degree

Kin1 Master ID	Kin1 archaeological context	Kin2 Master ID	Kin2 archaeological context	Mt Hg Kin1 - Kin2	Y Hg Ki1 - Kin2	Corrected kin coefficient	Estimated relatedness
DZS-1	Carpathian Basin Early Sarmatian Period	DZS-48	Carpathian Basin Early Sarmatian Period	U5a2a1 - H9a	XX - XX	0.076	3rd degree
TMH-199	Hungary_Avar_Late (Maróti et al 2022.)	TMHper496	Hungary_Avar (Unpublished)	HV+16311 - T2b	E1b1b1a1b1a6~ - E1b1b1a1b1a6a1a~	0.238	1st degree
TMH-199	Hungary_Avar_Late (Maróti et al 2022.)	SZODper829	Hungary_Avar (Unpublished)	HV+16311 - HV4a1	E1b1b1a1b1a6~ - E1b1b1a1b1a6a~	0.063	3rd degree
TMH-199	Hungary_Avar_Late (Maróti et al 2022.)	OFU-422	Carpathian Basin Hun Period	HV+16311 - HV4a1	E1b1b1a1b1a6~ - E1b1b1a1b1~	0.030	4th degree
SZODper829	Hungary_Avar (Unpublished)	TMHper496	Hungary_Avar (Unpublished)	HV4a1 - T2b	E1b1b1a1b1a6a~ - E1b1b1a1b1a6a1a~	0.026	4th degree
SZODper829	Hungary_Avar (Unpublished)	OFU-422	Carpathian Basin Hun Period	HV4a1 - HV4a1	E1b1b1a1b1a6a~ - E1b1b1a1b1~	0.011	uncertain
OFU-422	Carpathian Basin Hun Period	TMHper496	Hungary_Avar (Unpublished)	HV4a1 - T2b	E1b1b1a1b1~ - E1b1b1a1b1a6a1a~	0.002	uncertain
NKL-7	Carpathian Basin Sarmatian Unknown Period	TKDper15	Hungary_Avar (Unpublished)	X2b+226 - U2c1a	R1a1a1b1a1a1c1c1a1~ - R1a1a1b1a1a1c1c1a1~	0.030	4th degree
NKL-7	Carpathian Basin Sarmatian Unknown Period	TKDper33	Hungary_Avar (Unpublished)	X2b+226 - HV0+195	R1a1a1b1a1a1c1c1a1~ - R1a1a1b1a1a1c1c1a~	0.029	4th degree
NKL-7	Carpathian Basin Sarmatian Unknown Period	TKDper34	Hungary_Avar (Unpublished)	X2b+226 - U5a2d	R1a1a1b1a1a1c1c1a1~ - R1a1a1b1a1a1c1c1a1~	0.024	4th degree
NKL-7	Carpathian Basin Sarmatian Unknown Period	ULLper189	Hungary_Avar (Unpublished)	X2b+226 - U2c1a	R1a1a1b1a1a1c1c1a1~ - R1a1a1b1a1a1c1c1a~	0.029	4th degree
NKL-7	Carpathian Basin Sarmatian Unknown Period	TKDper18	Hungary_Avar (Unpublished)	X2b+226 - U5a2b	R1a1a1b1a1a1c1c1a1~ - R1a1a1b1a1a1c1~	0.011	uncertain
DZS-6	Carpathian Basin Early Sarmatian Period	DZS-43	Carpathian Basin Early Sarmatian Period	W3a1 - W3a1	XX - R1b1a1b1b3	0.033	4th degree
MM2-215	Carpathian Basin Late Sarmatian Period	MM2-245	Carpathian Basin Late Sarmatian Period	HV0+195 - HV	XX - XX	0.023	5th degree
TMH-509	Hungary_Avar_Late (Maróti et al 2022.)	OFU-168	Carpathian Basin Late Sarmatian Period	A8a1 - H1b1+16362	I1a3a1a2~ - I1a3a1~	0.043	4th degree
TMH-798	Hungary_Avar_Late (Maróti et al 2022.)	TMHper694	Hungary_Avar (Unpublished)	H3h - H16	I1a3a1a2~ - G1b	0.027	4th degree
OFU-168	Carpathian Basin Late Sarmatian Period	TMHper694	Hungary_Avar (Unpublished)	H1b1+16362 - H16	I1a3a1~ - G1b	0.042	4th degree

Kin1 Master ID	Kin1 archaeological context	Kin2 Master ID	Kin2 archaeological context	Mt Hg Kin1 - Kin2	Y Hg Ki1 - Kin2	Corrected kin coefficient	Estimated relatedness
TMH-798	Hungary_Avar_Late (Maróti et al 2022.)	OFU-168	Carpathian Basin Late Sarmatian Period	H3h - H1b1+16362	I1a3a1a2~ - I1a3a1~	0.020	5th degree
TMH-509	Hungary_Avar_Late (Maróti et al 2022.)	TMH-798	Hungary_Avar_Late (Maróti et al 2022.)	A8a1 - H3h	I1a3a1a2~ - I1a3a1a2~	0.000	uncertain
TMH-509	Hungary_Avar_Late (Maróti et al 2022.)	TMHper694	Hungary_Avar (Unpublished)	A8a1 - H16	I1a3a1a2~ - G1b	0.003	uncertain
OFU-190	Carpathian Basin Hun Period	ULLper64	Hungary_Avar (Unpublished)	T2a1a - T2e	XX - XX	0.021	5th degree
OFU-190	Carpathian Basin Hun Period	ULLperD	Hungary_Avar (Unpublished)	T2a1a - H11a5	XX - R1b1a1b1b3a	0.020	5th degree

Supplementary Table 5: Short label abbreviations of publicly available samples used in various analyses. The reference format was lifted from the Allen Ancient DNA Resource (AADR, Mallick et al., 2024). An "x" indicates the specific analyses in which each sample was used.

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
AUT_ROM	R10654	Austria_Klosterneuburg_Roman_oLevant.SG	AntoniobioRxiv2022				x
AUT_ROM	R10656	Austria_Klosterneuburg_Roman.SG	AntoniobioRxiv2022				x
AUT_ROM	R10657	Austria_Klosterneuburg_Roman.SG	AntoniobioRxiv2022				x
AUT_ROM	R10658	Austria_Klosterneuburg_Roman_oLevant_contam.SG	AntoniobioRxiv2022				x
AUT_ROM	R10659	Austria_Klosterneuburg_Roman.SG	AntoniobioRxiv2022				x
AUT_ROM	R10660	Austria_Klosterneuburg_Roman_oLevant.SG	AntoniobioRxiv2022				x
AUT_ROM	R10665	Austria_Ovilava_Roman.SG	AntoniobioRxiv2022				x
AUT_ROM	R10666	Austria_Ovilava_Roman.SG	AntoniobioRxiv2022				x
AUT_ROM	R10668	Austria_Ovilava_Roman_oLevant.SG	AntoniobioRxiv2022				x
AUT_ROM	R10670	Austria_Ovilava_Roman.SG	AntoniobioRxiv2022				x
CZE_BA_CWC	I6695	Czech_CordedWare	NarasimhanPattersonScience2019		x		
CZE_BA_CWC	I6696	Czech_CordedWare	NarasimhanPattersonScience2019		x		
CZE_BA_CWC	I7207	Czech_CordedWare	NarasimhanPattersonScience2019		x		
CZE_BA_CWC	I7208	Czech_CordedWare	NarasimhanPattersonScience2019		x		
CZE_BA_CWC	I7209	Czech_CordedWare	NarasimhanPattersonScience2019		x		
CZE_BA_CWC	I7279	Czech_CordedWare	OlaldeNature2018		x		
CZE_BA_CWC	I7280	Czech_CordedWare	OlaldeNature2018		x		
DEU_BA_CWC	I0049	Germany_CordedWare	MathiesonNature2015		x		
DEU_BA_CWC	I0103	Germany_CordedWare	MathiesonNature2015		x		
DEU_BA_CWC	I0104	Germany_CordedWare	MathiesonNature2015		x		
DEU_BA_CWC	I1532	Germany_CordedWare	MathiesonNature2015		x		
DEU_EMED_BAV	AED204	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
DEU_EMED_BAV	Alh1	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	Alh10	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	Alh3a	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	NW54	Germany_EarlyMedieval_o1.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR220c	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR228	Germany_EarlyMedieval_o3.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR300b	Germany_EarlyMedieval_o1.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR310	Germany_EarlyMedieval_o1.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR328c	Germany_EarlyMedieval_o2.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR355c	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR480	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_BAV	STR486	Germany_EarlyMedieval.SG	VeeramahPNAS2018				x
DEU_EMED_HASS	R11866	Germany_Hassleben_Germanic.SG	AntoniobioRxiv2022		x		
DEU_EMED_HASS	R11867	Germany_Hassleben_Germanic_ elite_1.SG	AntoniobioRxiv2022		x		x
DEU_EMED_HASS	R11868	Germany_Hassleben_Germanic.SG	AntoniobioRxiv2022		x		x
DEU_EMED_HASS	R11872	Germany_Hassleben_Germanic.SG	AntoniobioRxiv2022		x		x
DEU_EMED_HASS	R11873	Germany_Hassleben_Germanic.SG	AntoniobioRxiv2022		x		x
DEU_EMED_HASS	R11875	Germany_Hassleben_Germanic_ oNorthernEurope_contam.SG	AntoniobioRxiv2022				x
Ethiopia_4500BP	I5950	Ethiopia_4500BP.SG	LipsonSawchukNature2022			x	
GBR_BA_BB	I2443	England_BellBeaker	OlaldeNature2018		x		
GBR_BA_BB	I2445	England_BellBeaker	OlaldeNature2018		x		
GBR_BA_BB	I2447	England_BellBeaker	OlaldeNature2018		x		
GBR_BA_BB	I2450	England_BellBeaker	OlaldeNature2018		x		
GBR_BA_BB	I2454	England_BellBeaker_lowEEF	OlaldeNature2018		x		
GBR_BA_BB	I3255	England_BellBeaker	OlaldeNature2018		x		
GBR_BA_BB	I3256	England_BellBeaker	OlaldeNature2018		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
GBR_BA_BB	I6680	England_BellBeaker_mediumEEF	OlaldeNature2018		x		
GBR_BA_BB	I6774	England_BellBeaker_lowEEF	OlaldeNature2018		x		
GBR_BA_BB	I6775	England_BellBeaker	OlaldeNature2018		x		
GBR_BA_BB	I6778	England_BellBeaker_o	OlaldeNature2018		x		
GBR_EMED_SAX	I0157	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
GBR_EMED_SAX	I0159	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
GBR_EMED_SAX	I0161	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
GBR_EMED_SAX	I0769	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
GBR_EMED_SAX	I0773	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
GBR_EMED_SAX	I0774	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
GBR_EMED_SAX	I0777	England_EarlyMedieval_Saxon.SG	SchiffelsNatureCommunications2016				x
HRV_N_CAR	I3433	Croatia_N_Cardial	MathiesonNature2018		x		
HRV_N_CAR	I3947	Croatia_N_Cardial	MathiesonNature2018		x		
HRV_N_CAR	I3948	Croatia_N_Cardial	MathiesonNature2018		x		
HRV_ROM	POP23	Croatia_Popova_RomanP.SG	FreilichPinhasiScientificReports2021		x		
HRV_ROM	R2040	Croatia_SisakPogorelec_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2041	Croatia_SisakPogorelec_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2042	Croatia_SisakPogorelec_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2045	Croatia_Sipar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2050	Croatia_MirineFulfinum_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2051	Croatia_MirineFulfinum_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2053	Croatia_MirineFulfinum_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2055	Croatia_Metz_GalloRoman.SG	AntoniobioRxiv2022				x
HRV_ROM	R2057	Croatia_Metz_GalloRoman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2058	Croatia_Metz_GalloRoman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2065	Croatia_Metz_GalloRoman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R2066	Croatia_Metz_GalloRoman.SG	AntoniobioRxiv2022		x		x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HRV_ROM	R3542	Croatia_BeliManastir_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3543	Croatia_Tilurium_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3544	Croatia_Tilurium_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3545	Croatia_Tilurium_Roman_ oCaucasus.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3547	Croatia_NovoSeloBunje_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3655	Croatia_Mursa_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3656	Croatia_Mursa_Roman.SG	AntoniobioRxiv2022		x		
HRV_ROM	R3657	Croatia_Mursa_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3659	Croatia_Scitarjevo_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3660	Croatia_Scitarjevo_Roman_ oNorthEurope.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3662	Croatia_Sipar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3663	Croatia_Sipar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3664	Croatia_Sipar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3665	Croatia_Dragulin_Roman.SG	AntoniobioRxiv2022				x
HRV_ROM	R3670	Croatia_Policija_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3685	Croatia_Velic_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3742	Croatia_Zadar_Roman_oLevant.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3743	Croatia_Zadar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3744	Croatia_Zadar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3745	Croatia_Zadar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3746	Croatia_Zadar_Roman.SG	AntoniobioRxiv2022		x		x
HRV_ROM	R3747	Croatia_Zadar_Roman.SG	AntoniobioRxiv2022		x		x
HUN_AVAR_AC	AN-376	Hungary_Avar_Asia_Core_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	CS-465	Hungary_Avar_Asia_Core_ Early_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	CSPF-114	Hungary_Avar_Asia_Core_Early	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_AVAR_AC	CSPF-213	Hungary_Avar_Asia_Core_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	CSPF-37	Hungary_Avar_Asia_Core_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	FGD-4	Hungary_Avar_Asia_Core_ Early_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	FU-215	Hungary_Avar_Asia_Core_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	KFP-30a	Hungary_Avar_Asia_Core_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	KFP-31	Hungary_Avar_Asia_Core_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	KV-3369	Hungary_Avar_Asia_Core_ Early_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_AC	MM-245	Hungary_Avar_Asia_Core_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	ACG-19	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	AN-286	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	CSB-9	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	DK-701	Hungary_Avar_Early_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_E	FU-193	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	HC-168	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MM-131	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MM-151	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MM-240	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MM-61	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MM-80	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MM-83	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MS-43	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	MT-29	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SSD-17	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SSD-58	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZF-181	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZF-26	Hungary_Avar_Early	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_AVAR_E	SZF-371	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZK-102	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZK-213	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZOD1-127	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZOD1-187	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZOD1-554	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZOD1-76	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZOD1-829	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_E	SZRV-54	Hungary_Avar_Early	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ALT-369	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ALT-412	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ALT-414	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ALT-596	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-11	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-14	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-16	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-17	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-19	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-20	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-21	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-24	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-29	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-36	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-38	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-43	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-48	Hungary_Avar_Late	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_AVAR_L	ARK-49	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-50	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	ARK-6	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	HH-10	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	HH-22	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	JHT-130	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	JHT-30	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK1-251	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK1-252	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK1-368	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK1-541	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK2-441	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK2-445	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KK2-670	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	KV-3450	Hungary_Avar_Late_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBH-37	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBH-52	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBT-106	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBT-108	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBT-3	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBT-51	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	OBT-56	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	PV-12	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	PV-200	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	PV-205	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZK-130	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZKT-265	Hungary_Avar_Late	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_AVAR_L	SZKT-311	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZKT-62	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZKT-70	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZKT-89	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZM-332	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZRV-168	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZRV-212	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZRV-266	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZRV-277	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZRV-316	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	SZRV-67	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	TMH-1273	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	TMH-199	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	TMH-509	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	TMH-798	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	TTSZ-43	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_L	VPB-307	Hungary_Avar_Late_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_L	VPB-31	Hungary_Avar_Late	MarotiTorokCurrBio2022				x
HUN_AVAR_M	ALT-224	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	ALT-442	Hungary_Avar_MiddleLate	MarotiTorokCurrBio2022				x
HUN_AVAR_M	ALT-77	Hungary_Avar_EarlyMiddle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	ARK-41	Hungary_Avar_MiddleLate	MarotiTorokCurrBio2022				x
HUN_AVAR_M	CSPF-182	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	HH-102	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	JHT-154	Hungary_Avar_MiddleLate	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KD-16	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KD-29	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_AVAR_M	KDA-188	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KDA-485	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KDA-517	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KDA-520	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KK1-245	Hungary_Avar_MiddleLate	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KK2-429	Hungary_Avar_MiddleLate	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KPM-14	Hungary_Avar_EarlyMiddle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KPM-23	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KPM-27	Hungary_Avar_MiddleLate	MarotiTorokCurrBio2022				x
HUN_AVAR_M	KV-3367	Hungary_Avar_Middle_Elite	MarotiTorokCurrBio2022				x
HUN_AVAR_M	MS-50	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	MT-17	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	MT-23	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	MT-74	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	PV-116	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	PV-125	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SSD-144	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SSD-151	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SSD-198	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SSD-35	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZK-180	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZK-83	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZM-24	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZM-255	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZM-259	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZM-38	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	SZRV-147	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_AVAR_M	TMH-388	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_AVAR_M	TMH-756	Hungary_Avar_Middle	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	K2-29	Hungary_Conq_Asia_Core_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	KeF1-10936	Hungary_Conq_Asia_Core_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	LB-1432	Hungary_Conq_Asia_Core_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	SZA-154	Hungary_Conq_Asia_Core_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	SZAK-4	Hungary_Conq_Asia_Core_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	SZAK-7	Hungary_Conq_Asia_Core_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_AC	TCS-2	Hungary_Conq_Asia_Core_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-157	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-229	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-231	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-245	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-43	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-50	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-86	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	HMSZ-88	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	IBE-154	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	IBE-161	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	IBE-176	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	IBE-206	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	NTH-1	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	NTH-19	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	NTH-2	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	NTH-20	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_CONQ_COM	PLE-200	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	PLE-216	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	PLE-23	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	PLE-38	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	PLE-57	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-103	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-106	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-175	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-182	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-251	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-41	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SH-81	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZA-20	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZA-29	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZA-44	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZA-52	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZA-7	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZOD-376	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZOD-394	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZOD-426	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	SZOD-566	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	VPB-561	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_COM	VPB-588	Hungary_Conq_Commoner	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	IBE-106	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	IBE-107	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	IBE-90	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	MH-106	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_CONQ_EARP	MH-107	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	MH-153	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	MH-88	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	PLE-115	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	PLE-327	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	PLE-384	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	PLE-418	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	PLE-441	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	PLE-95	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	VPB-118	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_EARP	VPB-600	Hungary_Early_Arpadian	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	AGY-49	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	AGY-75	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	AGY-87	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	AGY-92	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	BK-2	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	CSU-11	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K1-10	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K1-3286	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K2-16	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K2-18	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K2-33	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K2-61	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K3-12	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	K3-13	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	KH-500	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	KH-596	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_CONQ_ELITE	MH1-14	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	MH1-4	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	NK-2	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SE-114	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SE-16	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SE-23	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SE-64	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SEO-3	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SEO-4	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SO-5	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SP-10	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SP-2	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	SP-9	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	TCS-18	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	TCS-5	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	VPB-167	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_CONQ_ELITE	VPB-310	Hungary_Conq_Elite	MarotiTorokCurrBio2022				x
HUN_HUN	ASZK-1	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	CSB-3	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	KMT-2785	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	MSG-1	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	SEI-1	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	SEI-5	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	SEI-6	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	SZLA-646	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_HUN	VZ-12673	Hungary_Hun	MarotiTorokCurrBio2022				x
HUN_IA_LT	I18110	Hungary_IA_LaTene	PattersonNature2021		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_IA_LT	I18182	Hungary_IA_LaTene_o	PattersonNature2021		x		
HUN_IA_LT	I18183	Hungary_IA_LaTene_o	PattersonNature2021		x		
HUN_IA_LT	I18220	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18226	Hungary_IA_LaTene_o	PattersonNature2021		x		
HUN_IA_LT	I18488	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18489	Hungary_LaTene	PattersonNature2021		x	x	
HUN_IA_LT	I18491	Hungary_LaTene	PattersonNature2021		x	x	
HUN_IA_LT	I18492	Hungary_LaTene	PattersonNature2021		x	x	
HUN_IA_LT	I18493	Hungary_LaTene	PattersonNature2021		x	x	
HUN_IA_LT	I18526	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18527	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18528	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18529	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18530	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18531	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18832	Hungary_IA_LaTene_oEast	PattersonNature2021		x		
HUN_IA_LT	I18834	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18837	Hungary_IA_LaTene_oWest	PattersonNature2021		x		
HUN_IA_LT	I18838	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18839	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I18840	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I20752	Hungary_IA_LaTene	HarneyCheronetGenomeResearch2021		x		
HUN_IA_LT	I25509	Hungary_IA_LaTene_o3	PattersonNature2021		x		
HUN_IA_LT	I25510	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I25512	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I25518	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I25519	Hungary_IA_LaTene	PattersonNature2021		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
HUN_IA_LT	I25522	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I25524	Hungary_IA_LaTene_o3	PattersonNature2021		x		
HUN_IA_LT	I4996	Hungary_IA_LaTene	PattersonNature2021		x		
HUN_IA_LT	I4998	Hungary_IA_LaTene_o	PattersonNature2021		x		
HUN_IA_SCY	DA191	Hungary_IA_Scythian.SG	DamgaardNature2018		x	x	x
HUN_IA_SCY	DA194	Hungary_IA_Scythian.SG	DamgaardNature2018		x	x	x
HUN_IA_SCY	DA195	Hungary_IA_Scythian_oAegean.SG	DamgaardNature2018		x		x
HUN_IA_SCY	DA197	Hungary_IA_Scythian.SG	DamgaardNature2018		x	x	x
HUN_IA_SCY	DA198	Hungary_IA_Scythian_oAegean.SG	DamgaardNature2018		x		x
HUN_IA_SCY	I20766	Hungary_IA_Scythian	HarneyCheronetGenomeResearch2021		x		
HUN_LN_LGY	I1495	Hungary_LN_Lengyel	MathiesonNature2015			x	
HUN_LN_LGY	I18691	Hungary_LN_Lengyel	PattersonNature2021			x	
HUN_LN_LGY	I1899	Hungary_LN_Lengyel	LipsonNature2017			x	
HUN_LN_LGY	I1900	Hungary_LN_Lengyel	LipsonNature2017			x	
HUN_LN_LGY	I1901	Hungary_LN_Lengyel	LipsonNature2017			x	
HUN_LN_LGY	I1902	Hungary_LN_Lengyel	LipsonNature2017			x	
HUN_LN_LGY	I1905	Hungary_LN_Lengyel	LipsonNature2017			x	
HUN_LN_LGY	I1906	Hungary_LN_Lengyel	LipsonNature2017			x	
HUN_LN_LGY	I2352	Hungary_LN_Lengyel	LipsonNature2017			x	
ITA_IA_BIV	R1554	Italy_Bivio_Roman.SG	AntoniobioRxiv2022				x
ITA_IA_BIV	R1555	Italy_Bivio_Roman.SG	AntoniobioRxiv2022				x
ITA_IA_BIV	R1556	Italy_Bivio_Roman.SG	AntoniobioRxiv2022				x
ITA_IA_BIV	R1557	Italy_Bivio_Roman.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R111	Italy_Imperial_o2.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R11109	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11112	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11113	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
ITA_IA_IMP	R11115	Italy_IsolaSacra_RomanImperial_oEurope.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11116	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11117	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11118	Italy_IsolaSacra_RomanImperial_oEurope.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11119	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11120	Italy_IsolaSacra_RomanImperial.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R11121	Italy_IsolaSacra_RomanImperial_oEurope.SG	AntoniobioRxiv2022				x
ITA_IA_IMP	R113	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R114	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R115	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R116	Italy_Imperial_oCentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R123	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R125	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R126	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R128	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R131	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R132	Italy_Imperial_oAfrica.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1543	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1544	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1545	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1547	Italy_Imperial_o4.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1548	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1549	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1550	Italy_Imperial_o4.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R1551	Italy_Imperial_o3.SG	AntonioGaoMootsScience2019				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
ITA_IA_IMP	R37	Italy_Imperial_oCentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R38	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R39	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R40	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R41	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R42	Italy_Imperial_o4.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R43	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R436	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R44	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R45	Italy_Imperial_o1.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R47	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R49	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R50	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R51	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R66	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R67	Italy_Imperial_o5.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R68	Italy_Imperial_o3.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R69	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R70	Italy_Imperial_o4.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R71	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R72	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R73	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R75	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R76	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R78	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R80	Italy_Imperial_o1.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R81	Italy_Imperial.SG	AntonioGaoMootsScience2019				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
ITA_IA_IMP	R835	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_IMP	R836	Italy_Imperial.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R1	Italy_IA_Republic_o.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R1015	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R1016	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R1021	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R435	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R437	Italy_IA_Republic_ oEasternMediterranean.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R473	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R474	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R475	Italy_IA_Republic_ oEasternMediterranean_o.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R850	Italy_IA_Republic_ oEasternMediterranean.SG	AntonioGaoMootsScience2019				x
ITA_IA_RP	R851	Italy_IA_Republic.SG	AntonioGaoMootsScience2019				x
ITA_LA	R105	Italy_LA_o1CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R106	Italy_LA_o2CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R107	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R108	Italy_LA_o1CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R109	Italy_LA_o1CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R110	Italy_LA_o1CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R117	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R118	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R120	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R121	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R122	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R130	Italy_LA.SG	AntonioGaoMootsScience2019				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
ITA_LA	R133	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R134	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R136	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R137	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R30	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R31	Italy_LA_o3CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R32	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R33	Italy_LA_o1CentralEuropean.SG	AntonioGaoMootsScience2019				x
ITA_LA	R34	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R35	Italy_LA.SG	AntonioGaoMootsScience2019				x
ITA_LA	R36	Italy_LA.SG	AntonioGaoMootsScience2019				x
KAZ_C_BOT	BOT14	Kazakhstan_Botai_Eneolithic	JeongNatureEcologyEvolution2019		x		
KAZ_C_BOT	BOT15	Kazakhstan_Botai_Eneolithic.SG	DamgaardScience2018		x		
KAZ_C_BOT	BOT2016	Kazakhstan_Botai_Eneolithic	JeongNatureEcologyEvolution2019		x		
KAZ_EIA_TAS	AKB001	Kazakhstan_Tasmola_EIA	GnecchiRusconeScienceAdvances2021		x		
KAZ_EIA_TAS	ESZ001	Kazakhstan_Tasmola_Saka_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_EIA_TAS	ESZ002	Kazakhstan_Tasmola_Saka_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_EIA_TAS	KKM001	Kazakhstan_Tasmola_EIA	GnecchiRusconeScienceAdvances2021		x		
KAZ_EIA_TAS	KSH002	Kazakhstan_Tasmola_EIA	GnecchiRusconeScienceAdvances2021		x		
KAZ_EIA_TAS	KZL004	Kazakhstan_Tasmola_LBA	GnecchiRusconeScienceAdvances2021		x		
KAZ_EIA_TAS	WAR001	Kazakhstan_Tasmola_EIA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE002	Kazakhstan_Berel_IA_o3	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE004	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE005	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE006	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE007	Kazakhstan_Berel_IA_o2	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE008	Kazakhstan_Berel_IA_o3	GnecchiRusconeScienceAdvances2021		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
KAZ_IA_BEREL	BRE009	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE010	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE012	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE013	Kazakhstan_Berel_IA	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	BRE014	Kazakhstan_Berel_IA_o3	GnecchiRusconeScienceAdvances2021		x		
KAZ_IA_BEREL	I0562	Kazakhstan_Berel_Pazyryk	UnterlanderNatureCommunications2017		x		
KAZ_IA_BEREL	I0563	Kazakhstan_Berel_Pazyryk	UnterlanderNatureCommunications2017		x		
KAZ_IA_CSAKA	DA10	Kazakhstan_CentralSaka_o1.SG	DamgaardNature2018		x		x
KAZ_IA_CSAKA	DA11	Kazakhstan_Central_Saka.SG	DamgaardNature2018		x		x
KAZ_IA_CSAKA	DA13	Kazakhstan_Central_Saka.SG	DamgaardNature2018		x		x
KAZ_IA_CSAKA	DA14	Kazakhstan_Central_Saka.SG	DamgaardNature2018				x
KAZ_IA_CSAKA	DA16	Kazakhstan_Central_Saka.SG	DamgaardNature2018		x		
KAZ_IA_CSAKA	DA17	Kazakhstan_CentralSaka_o2.SG	DamgaardNature2018				x
KAZ_IA_CSAKA	DA19	Kazakhstan_CentralSaka_o2.SG	DamgaardNature2018		x		x
KAZ_IA_CSAKA	I7110	Kazakhstan_IA_Saka_ dup.I7110.SG	DamgaardNature2018		x		
KAZ_IA_SARM	AIG001	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x		x	
KAZ_IA_SARM	AIG002	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x		x	
KAZ_IA_SARM	AIG003	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x	x		
KAZ_IA_SARM	AIG005	Kazakhstan_Sarmatian_IA_o4	GnecchiRusconeScienceAdvances2021	x			
KAZ_IA_SARM	AIG006	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x		x	
KAZ_IA_SARM	BSB001	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x	x	x	
KAZ_IA_SARM	BSB002	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x	x		
KAZ_IA_SARM	BSB003	Kazakhstan_Sarmatian_IA_ sister.BSB001	GnecchiRusconeScienceAdvances2021	x		x	
KAZ_IA_SARM	BSB004	Kazakhstan_Sarmatian_IA_o3	GnecchiRusconeScienceAdvances2021	x			
KAZ_IA_SARM	CLK001	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x		x	
KAZ_IA_SARM	DA202	Kazakhstan_Sarmatian.SG	DamgaardNature2018				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
KAZ_IA_SARM	DA26	Kazakhstan_Sarmatian.SG	DamgaardNature2018	x	x	x	x
KAZ_IA_SARM	DA30	Kazakhstan_Sarmatian.SG	DamgaardNature2018	x	x	x	
KAZ_IA_SARM	I11523	Kazakhstan_Sarmatian.SG	DamgaardNature2018	x			
KAZ_IA_SARM	I11537	Kazakhstan_EarlySarmatian	NarasimhanPattersonScience2019	x		x	
KAZ_IA_SARM	I11540	Kazakhstan_LateSarmatian	NarasimhanPattersonScience2019	x	x	x	
KAZ_IA_SARM	KBU001	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x	x	x	
KAZ_IA_SARM	KBU002	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x		x	
KAZ_IA_SARM	KBU003	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x			
KAZ_IA_SARM	KSK002	Kazakhstan_Sarmatian_IA_o1	GnecchiRusconeScienceAdvances2021	x	x	x	
KAZ_IA_SARM	SBL001	Kazakhstan_Sarmatian_IA_o2	GnecchiRusconeScienceAdvances2021	x	x		
KAZ_IA_SARM	SGZ001	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x	x		
KAZ_IA_SARM	SGZ002	Kazakhstan_Sarmatian_IA	GnecchiRusconeScienceAdvances2021	x	x		
KAZ_MLBA	I0507	Kazakhstan_MLBA_Dali	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I10110	Kazakhstan_Shoendykol_ MLBA_Fedorovo	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I10111	Kazakhstan_Shoendykol_ MLBA_Fedorovo	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I10112	Kazakhstan_Shoendykol_ MLBA_Fedorovo	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I10140	Kazakhstan_MLBA_Aktogai	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I1931	Kazakhstan_MLBA_Dali	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I3448	Kazakhstan_MLBA_Dali	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I3763	Kazakhstan_MLBA_Zevakinskiy	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I3767	Kazakhstan_AkMoustafa_MLBA1	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I3788	Kazakhstan_MLBA_OyDzhaylau	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I3861	Kazakhstan_MLBA_OyDzhaylau	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I3864	Kazakhstan_MLBA_Solyanka	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4262	Kazakhstan_MLBA_Karagash	NarasimhanPattersonScience2019		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
KAZ_MLBA	I4263	Kazakhstan_MLBA_Karagash	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4264	Kazakhstan_MLBA_Aktogai	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4265	Kazakhstan_MLBA_Aktogai	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4318	Kazakhstan_MLBA_Kairan	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4321	Kazakhstan_MLBA_KazakhMys	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4322	Kazakhstan_MLBA_KazakhMys	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4323	Kazakhstan_Kyzlbulak_MLBA1	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4566	Kazakhstan_MLBA_Kairan_o	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4773	Kazakhstan_MLBA_Aktogai	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4774	Kazakhstan_MLBA_Aktogai	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4776	Kazakhstan_MLBA_Kairan	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4779	Kazakhstan_MLBA_Kairan	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4782	Kazakhstan_MLBA_KazakhMys	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4783	Kazakhstan_MLBA_KazakhMys	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4784	Kazakhstan_MLBA_Kyzlbulak_o1	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4787	Kazakhstan_Taldysay_MLBA1	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4789	Kazakhstan_MLBA_OyDzhaylau	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4790	Kazakhstan_MLBA_OyDzhaylau	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I4794	Kazakhstan_Taldysay_MLBA2	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6707	Kazakhstan_Bylkyldak_MLBA	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6789	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6790	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6791	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6792	Kazakhstan_Maitan_MLBA_Alakul_o1	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6793	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6794	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
KAZ_MLBA	I6795	Kazakhstan_Maitan_MLBA_Alakul_o2	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6796	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6797	Kazakhstan_Maitan_MLBA_Alakul	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6799	Kazakhstan_MLBA_Alakul_Satan	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I6800	Kazakhstan_MLBA_Alakul_Lisakovskiy	NarasimhanPattersonScience2019		x		
KAZ_MLBA	I7060	Kazakhstan_MLBA_OyDzhaylau	NarasimhanPattersonScience2019		x		
KGZ_IA_TS_SAKA	DA49	Kyrgyzstan_TianShan_Saka.SG	DamgaardNature2018				x
KGZ_IA_TS_SAKA	DA50	Kyrgyzstan_TianShan_Saka.SG	DamgaardNature2018				x
KGZ_IA_TS_SAKA	DA53	Kyrgyzstan_TianShan_Saka_o2.SG	DamgaardNature2018				x
KGZ_IA_TS_SAKA	DA55	Kyrgyzstan_TianShan_Saka.SG	DamgaardNature2018				x
KGZ_IA_TS_SAKA	DA56	Kyrgyzstan_TianShan_Saka_o1.SG	DamgaardNature2018				x
LTU_ROM	R10830	Lithuania_Marvele_Roman.SG	AntoniobioRxiv2022				x
LTU_ROM	R10832	Lithuania_Marvele_Roman.SG	AntoniobioRxiv2022				x
LTU_ROM	R10836	Lithuania_Marvele_Roman.SG	AntoniobioRxiv2022				x
LTU_ROM	R10838	Lithuania_Marvele_Roman.SG	AntoniobioRxiv2022				x
MDA_IA_SCY	scy192	Moldova_Glinoe_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
MDA_IA_SCY	scy197	Moldova_Glinoe_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
MDA_IA_SCY	scy300	Moldova_Glinoe_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
MDA_IA_SCY	scy301	Moldova_Glinoe_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
MDA_IA_SCY	scy303	Moldova_Glinoe_Scythian_o2.SG	KrzewinskaScienceAdvances2018		x		x
MDA_IA_SCY	scy311	Moldova_Glinoe_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
MNE_ROM	R3478	Montenegro_Doclea_Roman.SG	AntoniobioRxiv2022				x
MNE_ROM	R3481	Montenegro_Doclea_Roman_oAegean.SG	AntoniobioRxiv2022				x
MNE_ROM	R3482	Montenegro_Doclea_Roman.SG	AntoniobioRxiv2022				x
MNE_ROM	R9918	Montenegro_Doclea_Roman.SG	AntoniobioRxiv2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
MNE_ROM	R9919	Montenegro_Doclea_Roman.SG	AntoniobioRxiv2022				x
MNE_ROM	R9920	Montenegro_Doclea_Roman.SG	AntoniobioRxiv2022				x
MNG_EIA_SG	I12960	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I12969	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I12971	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I13178	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I13963	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6349	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6352	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6353	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6357	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6359	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6365	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I6369	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
MNG_EIA_SG	I7032	Mongolia_EIA_SlabGrave_1	WangNature2021			x	
POL_LA_WC	R10618	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10620	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10625	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10626	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10631	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10633	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10634	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x
POL_LA_WC	R10636	Poland_Weklice_WielbarkCulture_Roman.SG	AntoniobioRxiv2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
POL_LA_WC	R11391	Poland_Weklice_WielbarkCulture_ Roman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10487	Portugal_Conimbriga_Roman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10488	Portugal_Conimbriga_Roman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10491	Portugal_MonteDaNora_ LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10494	Portugal_MonteDaNora_ LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10496	Portugal_MonteDaNora_ LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10499	Portugal_Miroico_LateRoman_ oAfrica.SG	AntoniobioRxiv2022				x
PRT_ROM	R10500	Portugal_Miroico_LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10501	Portugal_Miroico_LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10502	Portugal_Miroico_LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10503	Portugal_Miroico_LateRoman_ oAfrica.SG	AntoniobioRxiv2022				x
PRT_ROM	R10506	Portugal_Miroico_LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10507	Portugal_Miroico_LateRoman.SG	AntoniobioRxiv2022				x
PRT_ROM	R10508	Portugal_Miroico_LateRoman.SG	AntoniobioRxiv2022				x
ROU_EN	I2532	Romania_EN	MathiesonNature2018		x		
ROU_EN	I2533	Romania_EN	MathiesonNature2018		x		
RUS_BA_ANDR	RISE500	Russia_Andronovo.SG	AllentoftNature2015		x		
RUS_BA_ANDR	RISE503	Russia_Andronovo.SG	AllentoftNature2015		x		
RUS_BA_ANDR	RISE505	Russia_Andronovo.SG	AllentoftNature2015		x		
RUS_BA_ANDR	RISE512	Russia_Andronovo_o.SG	AllentoftNature2015		x		
RUS_EBA_YAM_STEP	I0231	Russia_Samara_EBA_Yamnaya	PattersonNature2021		x		
RUS_EBA_YAM_STEP	I0357	Russia_Samara_EBA_Yamnaya	MathiesonNature2015		x		
RUS_EBA_YAM_STEP	I0370	Russia_Samara_EBA_Yamnaya	MathiesonNature2015		x		
RUS_EBA_YAM_STEP	I0429	Russia_Samara_EBA_Yamnaya	MathiesonNature2015		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
RUS_EBA_YAM_STEP	I0438	Russia_Samara_EBA_Yamnaya	PattersonNature2021		x		
RUS_EBA_YAM_STEP	I0439	Russia_Samara_EBA_Yamnaya	MathiesonNature2015		x		
RUS_EBA_YAM_STEP	I0443	Russia_Samara_EBA_Yamnaya	MathiesonNature2015		x		
RUS_EBA_YAM_STEP	I0444	Russia_Samara_EBA_Yamnaya	MathiesonNature2015		x		
RUS_EBA_YAM_STEP	I7489	Russia_Samara_EBA_Yamnaya	NarasimhanPattersonScience2019		x		
RUS_EBA_YAM_STEP	RISE548	Russia_Kalmykia_EBA_ Yamnaya.SG	AllentoftNature2015		x		
RUS_EBA_YAM_STEP	RISE550	Russia_Kalmykia_EBA_ Yamnaya.SG	AllentoftNature2015		x		
RUS_EBA_YAM_STEP	RISE552	Russia_Kalmykia_EBA_ Yamnaya.SG	AllentoftNature2015		x		
RUS_HG	I0061	Russia_Karelia_HG.SG	FuNature2016		x		
RUS_HG	I0124	Russia_Samara_HG	MathiesonNature2015		x		
RUS_HG	I1960	Russia_Tyumen_HG	NarasimhanPattersonScience2019		x		
RUS_HG	I5766	Russia_Sosnoviy_HG	NarasimhanPattersonScience2019		x		
RUS_HG	I8743	Russia_Sidelkino_HG.SG	DamgaardScience2018		x		
RUS_HG	MA1	Russia_MA1_HG.SG	RaghavanNature2013		x		
RUS_HG	Ust_Ishim	Russia_Ust_Ishim_HG.DG	FuNature2014		x		
RUS_IA_SARM	DA134	Russia_Sarmatian.SG	DamgaardNature2018	x	x	x	x
RUS_IA_SARM	DA136	Russia_Sarmatian.SG	DamgaardNature2018	x	x		x
RUS_IA_SARM	DA139	Russia_Sarmatian.SG	DamgaardNature2018	x	x	x	x
RUS_IA_SARM	DA141	Russia_Sarmatian.SG	DamgaardNature2018	x	x	x	x
RUS_IA_SARM	DA143	Russia_Sarmatian.SG	DamgaardNature2018	x	x	x	
RUS_IA_SARM	DA144	Russia_Sarmatian.SG	DamgaardNature2018	x	x	x	x
RUS_IA_SARM	DA145	Russia_Sarmatian.SG	DamgaardNature2018	x		x	
RUS_IA_SARM	MJ-38	Russia_Sarmatian.SG	JarveCurrentBiology2019	x			
RUS_IA_SARM_E	I0574	Russia_Ural_EarlySarmatian	UnterlanderNatureCommunications2017	x	x		
RUS_IA_SARM_E	I0575	Russia_Ural_EarlySarmatian	UnterlanderNatureCommunications2017	x	x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
RUS_IA_SARM_E	LS-13	Russia_EarlySarmatian_ SouthernUrals.SG	JarveCurrentBiology2019	x	x		x
RUS_IA_SARM_E	MJ-39	Russia_EarlySarmatian_ SouthernUrals.SG	JarveCurrentBiology2019	x		x	
RUS_IA_SARM_E	MJ-41	Russia_EarlySarmatian_ SouthernUrals.SG	JarveCurrentBiology2019	x	x	x	x
RUS_IA_SARM_E	MJ-43	Russia_EarlySarmatian_ SouthernUrals.SG	JarveCurrentBiology2019	x	x		x
RUS_IA_SARM_E	MJ-56	Russia_EarlySarmatian_ SouthernUrals.SG	JarveCurrentBiology2019	x			x
RUS_IA_SARM_E	Pr10	Russia_EarlySarmatian.SG	VeeramahPNAS2018	x			
RUS_IA_SARM_E	Pr4	Russia_EarlySarmatian.SG	VeeramahPNAS2018	x		x	
RUS_IA_SARM_L	chy001	Russia_LateSarmatian.SG	KrzewinskaScienceAdvances2018	x	x	x	x
RUS_IA_SARM_L	chy002	Russia_LateSarmatian.SG	KrzewinskaScienceAdvances2018	x	x		x
RUS_IA_SARM_L	MJ-44	Russia_MiddleSarmatian_ SouthernUrals.SG	JarveCurrentBiology2019	x	x	x	
RUS_IA_SARM_L	tem001	Russia_LateSarmatian.SG	KrzewinskaScienceAdvances2018	x		x	
RUS_IA_SARM_L	tem002	Russia_LateSarmatian.SG	KrzewinskaScienceAdvances2018	x	x	x	x
RUS_IA_SARM_L	tem003	Russia_LateSarmatian.SG	KrzewinskaScienceAdvances2018	x	x		x
RUS_IA_SCY	MJ-42	Russia_EasternScythian_ SouthernUrals.SG	JarveCurrentBiology2019				x
RUS_LBA_SRU	I0232	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0234	Russia_Srubnaya	MathiesonNature2015		x		
RUS_LBA_SRU	I0235	Russia_Srubnaya	MathiesonNature2015		x		
RUS_LBA_SRU	I0354	Russia_Srubnaya_o1	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0358	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0359	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0361	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0422	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0423	Russia_Srubnaya	MathiesonNature2015		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
RUS_LBA_SRU	I0424	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_LBA_SRU	I0430	Russia_Srubnaya	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I1821	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I1851	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I1852	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I1856	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3389	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3390	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3391	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3392	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3394	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3395	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I3396	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I6716	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I6717	Russia_MLBA_Krasnoyarsk_o	NarasimhanPattersonScience2019		x		
RUS_MLBA_KRA	I6718	Russia_MLBA_Krasnoyarsk	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I0939	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I0943	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I0983	Russia_MLBA_Sintashta_o1	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I0984	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1008	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1011	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1018	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1020	Russia_MLBA_Sintashta_o2	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1024	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1027	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1028	Russia_MLBA_Sintashta_o3	NarasimhanPattersonScience2019		x		

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
RUS_MLBA_SINTH	I1029	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1053	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1057	Russia_MLBA_Sintashta_o2	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1063	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1064	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1082	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1089	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I1090	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	I7480	Russia_MLBA_Sintashta	NarasimhanPattersonScience2019		x		
RUS_MLBA_SINTH	RISE386	Russia_MLBA_Sintashta.SG	AllentoftNature2015		x		
RUS_MLBA_SINTH	RISE394	Russia_MLBA_Sintashta.SG	AllentoftNature2015		x		
RUS_MLBA_SINTH	RISE395	Russia_MLBA_Sintashta.SG	AllentoftNature2015		x		
RUS_N_DEV	NEO236	Russia_DevilsCave_N.SG	SikoraNature2019		x		
RUS_N_DEV	NEO238	Russia_DevilsCave_N.SG	SikoraNature2019		x		
RUS_N_DEV	NEO240	Russia_DevilsCave_N.SG	SikoraNature2019		x		
SRB_HUN	Vim2b	Serbia_Medieval_Gepidian.SG	VeeramahPNAS2018				x
SRB_ROM	R3918	Serbia_Sirmium_LateRoman.SG	AntoniobioRxiv2022				x
SRB_ROM	R3931	Serbia_Viminacium_Roman_ elite_1.SG	AntoniobioRxiv2022				x
SRB_ROM	R6693	Serbia_SvilosKrussevlje_Roman.SG	AntoniobioRxiv2022				x
SRB_ROM	R6701	Serbia_SvilosKrussevlje_Roman.SG	AntoniobioRxiv2022				x
SRB_ROM	R6730	Serbia_Sirmium_Roman.SG	AntoniobioRxiv2022				x
SRB_ROM	R6750	Serbia_Viminacium_Roman_ elite_2.SG	AntoniobioRxiv2022				x
SRB_ROM	R6756	Serbia_Viminacium_Roman_ elite_3.SG	AntoniobioRxiv2022				x
SRB_ROM	R6759	Serbia_Viminacium_Roman.SG	AntoniobioRxiv2022				x
SRB_ROM	R6764	Serbia_Naissus_LateAntiquity.SG	AntoniobioRxiv2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
SRB_ROM	R6769	Serbia_Naissus_LateAntiquity_ oLevant.SG	AntoniobioRxiv2022				x
SRB_ROM	R9669	Serbia_Viminacium_Roman_ elite_3.SG	AntoniobioRxiv2022				x
SRB_ROM	R9673	Serbia_Viminacium_Roman_ elite_1.SG	AntoniobioRxiv2022				x
SRB_ROM	R9674	Serbia_Viminacium_Roman_ elite_1.SG	AntoniobioRxiv2022				x
SVK_LA_GERM	R2206	Slovakia_TesarkeMlynany_ Germanic_MigrationPeriod.SG	AntoniobioRxiv2022				x
SVK_LA_GERM	R2207	Slovakia_TesarkeMlynany_ Germanic_MigrationPeriod.SG	AntoniobioRxiv2022				x
SVK_LA_GERM	R2208	Slovakia_TesarkeMlynany_ Germanic_MigrationPeriod.SG	AntoniobioRxiv2022				x
SVK_LA_GERM	R2209	Slovakia_TesarkeMlynany_ Germanic_MigrationPeriod.SG	AntoniobioRxiv2022				x
SVK_LA_GERM	R2210	Slovakia_TesarkeMlynany_ Germanic_MigrationPeriod.SG	AntoniobioRxiv2022				x
SVK_LA_GERM	R2211	Slovakia_TesarkeMlynany_ Germanic_MigrationPeriod.SG	AntoniobioRxiv2022				x
SVK_LA_POPR	DA119	Slovakia_Poprad.SG	DamgaardNature2018				x
SVK_ROM	R2200	Slovakia_BytcaHrabove_Puchov_ LaTene_Roman.SG	AntoniobioRxiv2022				x
SVK_ROM	R2201	Slovakia_BytcaHrabove_Puchov_ LaTene_Roman.SG	AntoniobioRxiv2022				x
SVK_ROM	R2202	Slovakia_Mikusovce_ LaTene_Roman.SG	AntoniobioRxiv2022				x
SVK_ROM	R2204	Slovakia_Zohor_Germanic_ Roman.SG	AntoniobioRxiv2022				x
SVN_ROM	R10467	Slovenia_Emona_Roman.SG	AntoniobioRxiv2022				x
SVN_ROM	R10469	Slovenia_Emona_Roman.SG	AntoniobioRxiv2022				x
SVN_ROM	R10471	Slovenia_Emona_Roman.SG	AntoniobioRxiv2022				x
SVN_ROM	R10473	Slovenia_Emona_Roman.SG	AntoniobioRxiv2022				x

Short Label	Master ID	Group Label	Reference AADR	PCA	ADMIXTURE	F4 analysis	IBD analysis
SVN_ROM	R10474	Slovenia_Emona_Roman_ oEurope.SG	AntoniobioRxiv2022				x
SVN_ROM	R10477	Slovenia_Emona_Roman.SG	AntoniobioRxiv2022				x
SVN_ROM	R10478	Slovenia_Emona_Roman.SG	AntoniobioRxiv2022				x
UKR_IA_CIMM	MJ-12	Ukraine_Cimmerians.SG	JarveCurrentBiology2019				x
UKR_IA_SCY	MJ-16	Ukraine_IA_WesternScythian.SG	JarveCurrentBiology2019				x
UKR_IA_SCY	MJ-34	Ukraine_IA_WesternScythian.SG	JarveCurrentBiology2019		x		x
UKR_IA_SCY	scy009	Ukraine_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
UKR_IA_SCY	scy010	Ukraine_Scythian.SG	KrzewinskaScienceAdvances2018		x		x
UKR_IA_SCY	scy011	Ukraine_Scythian.SG	KrzewinskaScienceAdvances2018				x
WHG	Bichon	Switzerland_Bichon.SG	JonesNatureCommunications2015		x		
WHG	I0001	Luxembourg_Loschbour.DG	LazaridisNature2014		x		
WHG	I26770	Italy_Mesolithic.SG	AntonioGaoMootsScience2019		x		
WHG	I26771	Italy_Mesolithic.SG	AntonioGaoMootsScience2019		x		
WHG	I26772	Italy_Mesolithic.SG	AntonioGaoMootsScience2019		x		
WHG	I6767	England_Mesolithic.SG	BraceDiekmannNatureEcologyEvolution2019		x		
WHG	Ranchot88	France_Ranchot88	FuNature2016		x		
WHG	Spiginas4	Lithuania_Mesolithic	MittnikNatureCommunications2018		x		
WHG	Villabruna	Italy_North_Villabruna_HG	FuNature2016		x		