bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tracing genetic connections of ancient Hungarians to the 6-14th century populations of the Volga-Ural region

Bea Szeifert[1,2], Dániel Gerber[1,2], Veronika Csáky[1], Péter Langó[3,15], Dmitrii Alekseevich Stashenkov[4], Aleksandr Aleksandrovich Khokhlov[5], Ayrat Gabitovich Sitdikov[6], Ilgizar Ravil'evich Gazimzyanov[6], Elizaveta Valer'evna Volkova[6], Natalia Petrovna Matveeva[7], Alexander Sergejevich Zelenkov[7], Olga Evgen'evna Poshekhonova[8], Anastasiia Viktorovna Sleptsova[8], Konstantin Gennadievich Karacharov[9], Viktoria Vladimirovna Ilyushina[8], Boris Aleksandrovich Konikov[10], Flarit Abdulhaevich Sungatov[11], Alexander Gennadievich Kolonskikh[12], Sergei Gennad'evich Botalov[13], Ivan Valer'evich Grudochko[13], Balázs Egyed[14], Balázs Gusztáv Mende[1], Attila Türk[15,16], Anna Szécsényi-Nagy[1]

[1]Institute of Archaeogenomics, Research Centre for the Humanities, Eötvös Loránd Research Network (ELKH), Budapest, Hungary
[2]Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/C. 1117 Budapest, Hungary
[3]Institute of Archaeology, Research Centre for the Humanities, Eötvös Loránd Research Network (ELKH), Budapest, Hungary
[4]Samara Regional Museum of History and Local Lore named after P. V. Alabina, Samara, Russia
[5]Samara State University of Social Sciences and Education, Department of Biology, Ecology and Teaching Methods Samara, Russia
[6]Kazan Federal University, 420111, Russia, Kazan, Pushkina str. 1/55
[7]University of Tyumen, Tyumen, Russia
[8]Tyumen Scientific Centre SB RAS, Institute of the Problems of Northern Development, Tyumen, Russia
[9]Self-employed entrepreneur, Russia, Yekaterinburg
[10]Omsk Popov Production Association Russia, Omsk, 10 let Oktzabrya st, 195, 644009.
[11]Institute of History, Language and Literature of Scientific Center in Ufa of Russian Academy of Science, Ufa, Russia
[12]Institute of Ethnological Studies of R.G. Kuzeev, Ufa Scientific Center, Russian Academy of Sciences, Ufa, Russia
[13]South Ural State University, Chelyabinsk, Russia
[14]Department of Genetics, ELTE Eötvös Loránd University, Budapest, Hungary
[15]Institute of Archaeological Sciences, Faculty of Humanities and Social Sciences, Pázmány Péter Catholic University, Budapest, Hungary
[16]Hungarian Prehistory Researchgroup, Research Centre for the Humanities, Eötvös Loránd Research Network (ELKH), Budapest, Hungary

Correspondence to: szeifert.bea@abtk.hu, turk.attila@abtk.hu, szecsenyi-nagy.anna@abtk.hu

Keywords: Volga-Ural region, archaeogenetics, population genetics, phylogenetics, early history of Hungarians

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Most of the early Hungarian tribes originated from the Volga-Kama and South-Ural regions, where they were composed of a mixed population based on historical, philological and archaeological data. Here we present the uniparental genetic makeup of the medieval era of these regions that served as a melting pot for ethnic groups with different linguistic and historical backgrounds. Representing diverse cultural contexts, the new genetic data originates from ancient proto-Ob-Ugric people from Western Siberia (6-13th century), pre-Conquest period and subsisting Hungarians from Volga-Ural region (6-14th century) and their neighbours. By examining the Hungarian prehistory's eastern archaeological traits, we also study their genetic composition and origin in an interdisciplinary framework.

We analysed 110 deep sequenced mitogenomes and 42 Y-chromosome haplotypes from 18 archaeological sites from Russia. The results support the studied groups' genetic relationships regardless of geographical distances, suggesting large-scale mobility. We detected long-lasting genetic connections between the sites representing the Kushnarenkovo and Chiyalik cultures and the Carpathian Basin Hungarians and confirmed the Uralic transmission of several East-Eurasian uniparental lineages in their genepool. Based on phylogenetics, we demonstrate and model the connections and splits of the studied Volga-Ural and conqueror groups.

Early Hungarians and their alliances conquested the Carpathian Basin in the 890s AD. Re-analysis of the Hungarian conquerors' maternal genepool reveals numerous surviving maternal relationships in both sexes, therefore we conclude that men and women came to the Carpathian Basin together, and although they were subsequently genetically fused into the local population of the area, certain eastern lineages survived for centuries.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The Hungarians are the sole Uralic-speaking people in Central Europe today. The earliest known settlement area that can be associated with their ancestors is the territory bordered by the Rivers Tobol, Irtysh, and Ishim in the Trans-Urals and the western zone of south-western Siberia. Their artefacts appear among the finds left by the descendants of the Iron Age Sargat cultures, on both sides of the Urals, mainly in the distribution areas of the early medieval Bakal and Potchevash cultures (1) (2) (3) (4). In the 6-8th centuries the early Hungarians (together with other groups) lived most probably in the southern Ural region, and their archaeological remains were a substantial part of the Kushnarenkovo-Karayakupovo culture (5) (6) (2). At the beginning of the 9th century AD, the ancestors of the Hungarians crossed the River Volga and moved to the territory lying to the north of the Black Sea (Subbotsi-type sites), where they became the neighbours of the Khazars and Slavic-speaking peoples (7) (2) (8) (9). Later, leaving the Khazar Khaganate along with Kabars, they settled in the Carpathian Basin (CB) in the 890s (10) (6) (2) (9). Meanwhile, Hungarians who remained in the Volga–Ural region were reported in the middle of the 13th century (11) (12), whose tangible heritage was associated with the Chiyalik culture in archaeological research, the area of which is known as *Magna Hungaria* (5) (13) (14) (15) (16). Historical and linguistic data suggesting that one part of the Hungarians conquering the Carpathian Basin came from the Southern Urals and Trans-Urals and that is also supported by the findings of archaeological research (6) (2) (8) (17) (9).

The Volga–Ural region was subjected to several waves of migration, and its population was extremely complex in both historical and genetic terms. There is little genetic information about the medieval period of Central Eurasia but studies on its Iron and Bronze Age populations (18) (19) (20) serve as important reference points. The populations living in the area today have been mainly studied genetically in the context of the Uralic language family. In most cases, similarities could be pointed out among the groups belonging to the Uralic language family living in the Volga-Ural region. Furthermore, connections could also be detected with their geographical neighbours, the Bashkir, Chuvash and Tatar groups (21) (22). In the case of the Kushnarenkovo culture (from the Trans-Uralic Uyelgi site) as well as those representing the Lomovatovo and Nevolino cultures associated populations (Cis-Ural is the western foreland of the Urals), also show extensive genetic connections to the conquering Hungarians (23).

The composition of the uniparental genetic lines (24) (25) (26) (27) of the Hungarians living today in the CB is similar to that of other European peoples (21), but the maternal (23) (28) (29) (30) (31) (32) (33) (34) and paternal (33) (35) (36) lineages of the population of the archaeological Conquest period in CB in the late 9th-10th centuries (henceforth: conqueror Hungarians, in short conquerors) show a different picture. Their maternal lineages are similar to modern Tatars, in addition with a significant Eastern Eurasian component. On this basis relationships with Potapovka, Poltavka, and Srubnaya cultures associated populations were suggested, and connections with the Scythians and Huns were raised (29). Due to the great time gap, however, the connections with these groups are rather indirect. Conqueror paternal lineage composition is most similar to today's Bashkirs (35)(36). Fóthi et al. traced back the origins of these paternal lineages to three areas between the Lake Baikal and the Altai Mountains, between Western Siberia and the Southern Urals, and to the regions of the Black Sea and the North Caucasus (35), but due to deep genetic divergence dates firm conclusions on tribal origins can not be drawn from this observation.

Previous studies on the 10th to 12th century CB population divided individuals into two groups of "elite" and "common" people based on the funerary furnishing and burial customs (28) (29) (30) (31) (32) (33) (34) (35) (36). These concepts have become archaeologically incomprehensible and vague by today, therefore a division with a well-defined description focusing on the number of graves discovered in the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cemeteries and the chronology of their usage was introduced (37) (38). Our analyses adopted this approach by dividing the conqueror dataset into three main groups based on Kovács(37): KL-IV (largely correspondent to former group "elite"), this group characterized by 10th century small cemeteries of the camps that were the primary settlements of the nomadic Hungarians; KL-V and KL-VI (largely correspondent to former group "commoners"). KL-V represents 10th century cemeteries of villages with a large number of burials and KL-VI contains large village cemeteries that were opened in the 10th century and used until the 11th and 12th centuries.

Our study focuses on regions in present-day Russia that were important for the genesis of several Turkic (e.g. Bashkirs, Tatars) and Uralic speaking groups, such as the Maris, Khantys and the Hungarians as well. We analyse the uniparental markers of 112 individuals (6th–14th centuries) representing 13 sites located in the region of the Volga and Southern Urals (collectively Volga-Ural region or in short VUR), as well as five sites associated with early Ob-Ugric people (6-13th) living in Western Siberia (for more information about the studied sites see Fig. 1, Methods and Supplementary Material Chapter A and Table S1). Our objective has been to characterise genetically the cultural and ethnic hub of the VUR, focusing on those sites that can be linked to the early Hungarians archaeologically or their neighbours in a geographical sense, and to determine whether the possible ancestors of the Hungarians show biological link to their neighbours and the conquerors of the CB. We also examined how much of the genetic makeup of the VUR was preserved by the population of the CB between the 10th and the 12th centuries.

In our study we analyse 10 groups (Table 1) which were formed according to archaeological, chronological and geographic aspects. In addition to the investigated sites, we used database data of the conquerors from the Carpathian Basin (23) (28) (29) (30) (KL-IV–VI), the Cis-Ural group (Bayanovo, Brody, Bartim, Sukhoy Log) and the Uyelgi site (23).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

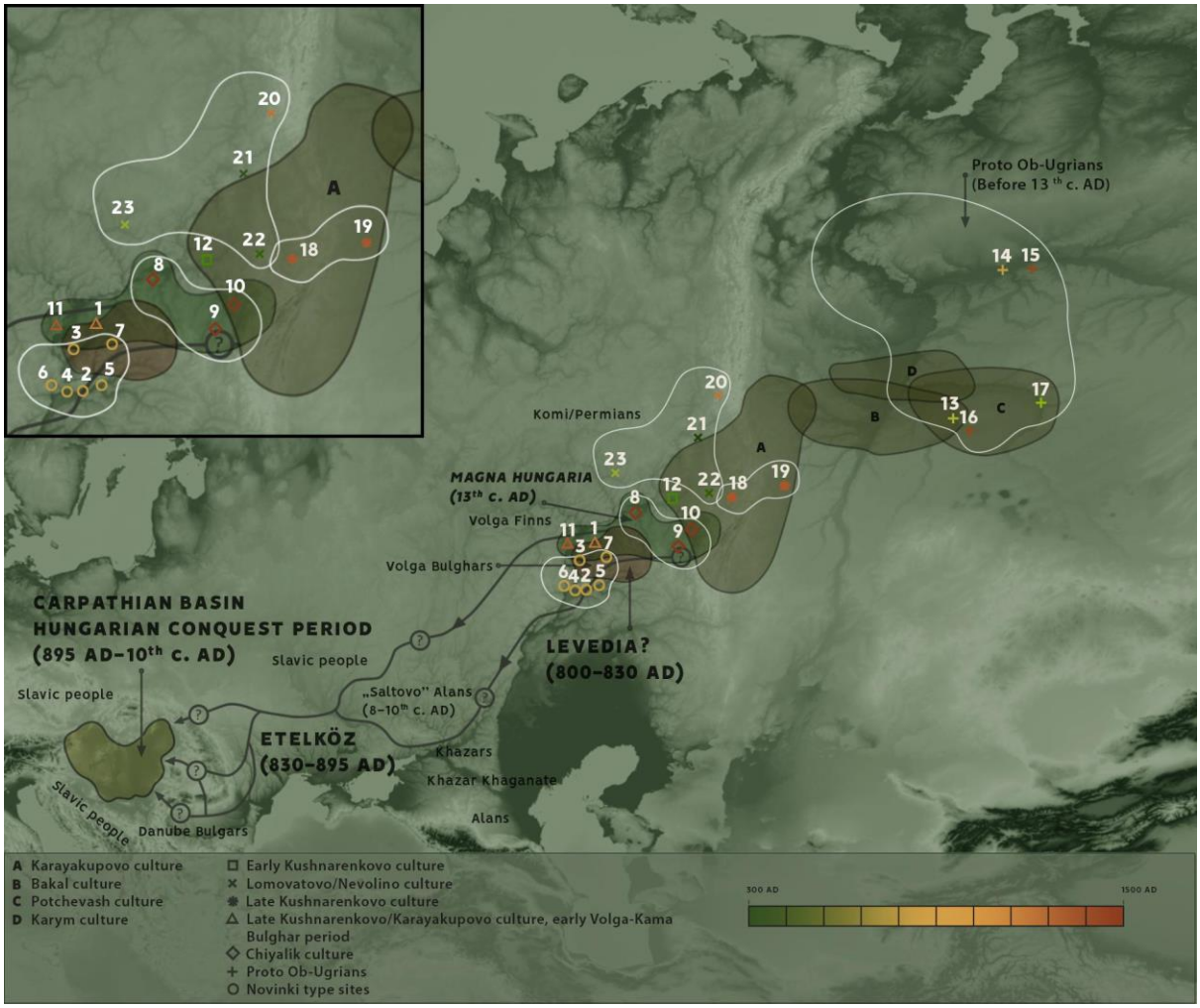

Figure 1: The supposed migration route of the early Hungarians (arrows) and the regions which could be linked to them (uppercase letters). The sites were grouped according to archaeological and chronological aspects; the formed groups are marked with white outlines. The investigated sites and the groups formed from them: Bolshie Tigani (1); Novinki group: Novinki (2), Mulovka (3), Brusyany (4), Lebyazhinka (5), Malaya Ryazan (6), Shilovka (7); Chiyalik group: Gulyukovo (8), Novo Hozyatovo (9), Gornovo (10); Tankeevka (11); Bustanaevo(12); Proto-Ob-Ugric group: Vikulovo (13), Barshov Gorodok (14), Ivanov Mis (15), Panovo (16), Ust-Tara (17); Uyelgi+Karanayevo group: Karanayevo (18), Uyelgi (19); Cis-Ural group: Bayanovo (20), Brody (21), Bartim (22), Sukhoy Log (23) (source of 19-23.: (Csáky et al 2020 (23)).
Source of map: Qgis v3.16.0 Topographic WMS - by terrestris (https://ows.terrestris.de/osm/service?).
Modifications were made with Adobe Photoshop CS6 and Adobe Photoshop 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results and discussion

Primary observations and paternal lineages

73 different haplotypes could be detected based on the complete mitochondrial genome sequences obtained from 112 newly examined individuals (Supplementary Material, Table S2). These belong to 15 macro-haplogroups (A, B, C, D, F, G, H, J, K, M, N, T, U, V, Z), which – with the exception of Z – were also described in the conquerors (Supplementary Material Table S7). Other macrohaplogroups (I, W, X, Y) are also present in conqueror groups, which may have reached them from other sources or could not be detected in the studied eastern groups due to their small sample sizes. Eastern-western geographical distribution of mitochondrial macrohaplogroups' frequency observable today is also reflected in our groups, as e.g. proto-Ob-Ugric associated individuals mostly possess eastern-type (A, B, C, D, G, M), and e.g. Tankeevka mostly western-type (H, T, U) lineages, see Fig. 2, Supplementary Material Table S2 and Table S7.

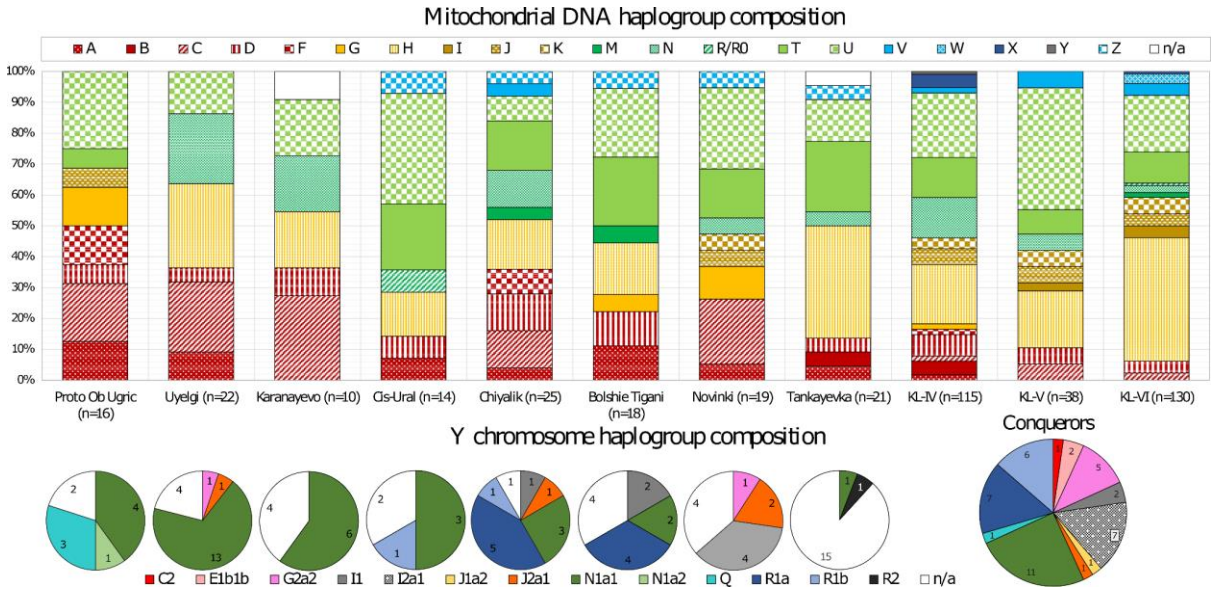

Figure 2: The mitochondrial (bar graph) and Y-chromosome (pie charts) haplogroup compositions of the analysed groups. In the case of the Hungarian conquerors, most of the Y-chromosomal data are known from the KL-IV group (35) (36). Due to the underrepresentation of KL-V and KL-VI groups, the conquerors were merged into one group in this figure.

For 42 males out of 72 yielded identifiable Y-chromosome haplotypes, see Methods and Supplementary Material, Table S2 and Table S10. The paternal lineages of the examined individuals can be classified to eight major haplogroups (HGs) (G2a2, I1, J2a1, N1a1, N1a2, Q, R1a, R1b). Most of them also appear at several sites associated with the Hungarians, as well as among the Hungarian conquerors (Fig. 2). The Q-L330 subgroup (ISOGG 15.73 Q1b1a3) was detected only in the proto-Ob-Ugric group, which corresponds to its Altai or Siberian origin (39). HG N1a1-M46 is present at all sites associated with the early Hungarians and among the conquerors, but could not be detected in the Novinki group.
36.8% of the known paternal lineages of the Hungarian conquerors belong to the HG N, although on average it is rare (1–2%) among today's Hungarians compared to other members of the Uralic language family (21) (40) (41). Interestingly, this proportion is somewhat higher in geographically isolated regions in e.g Eastern-Hungary Bodrogköz region, reaching a frequency of 6.1% (42). Within N1a1-M46, the N1a-Z1936 (ISOGG v15.73 N1a1a1a1a2) subclade potentially represents relationship between the representatives of the Uralic language family, which subgroup has also been described in the conquerors (35) (36).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Based on the STR network seen on Fig. 3, the individuals from Uyelgi+Karanayevo (Karanayevo) and Chiyalik groups (Gornovo) belong to one of the subgroups of N1a1a1a1a2a1c (ISOGG v15.73; SNP: B539/PH3340) fitting in the genetic composition of the Volga–Ural region, and they are clustered together with the samples from Uyelgi, and present-day Khantys, Mansis and Hungarians (all belong to the Ob-Ugric branch of Uralic languages), as well as with Bashkirs and Tatars the Volga–Ural region (23) (27). Because the Y-SNP results from literature and the Y-STR networks are not always compatible, we cannot define more downstream SNPs and subgroups, then N1a-B539/PH3340.
On R1a networks samples from Novinki and Bolshie Tigani groups (Supplementary Material, Fig. S40-S41) are related to samples from the Middle East, while individuals from the Chiyalik group are close to Russian and Belorussian males. In addition, these analyses also show intra-site connections. Based on the Y-STR patterns, another intra-site link was detected between two males from Bolshie Tigani sharing HG I1.

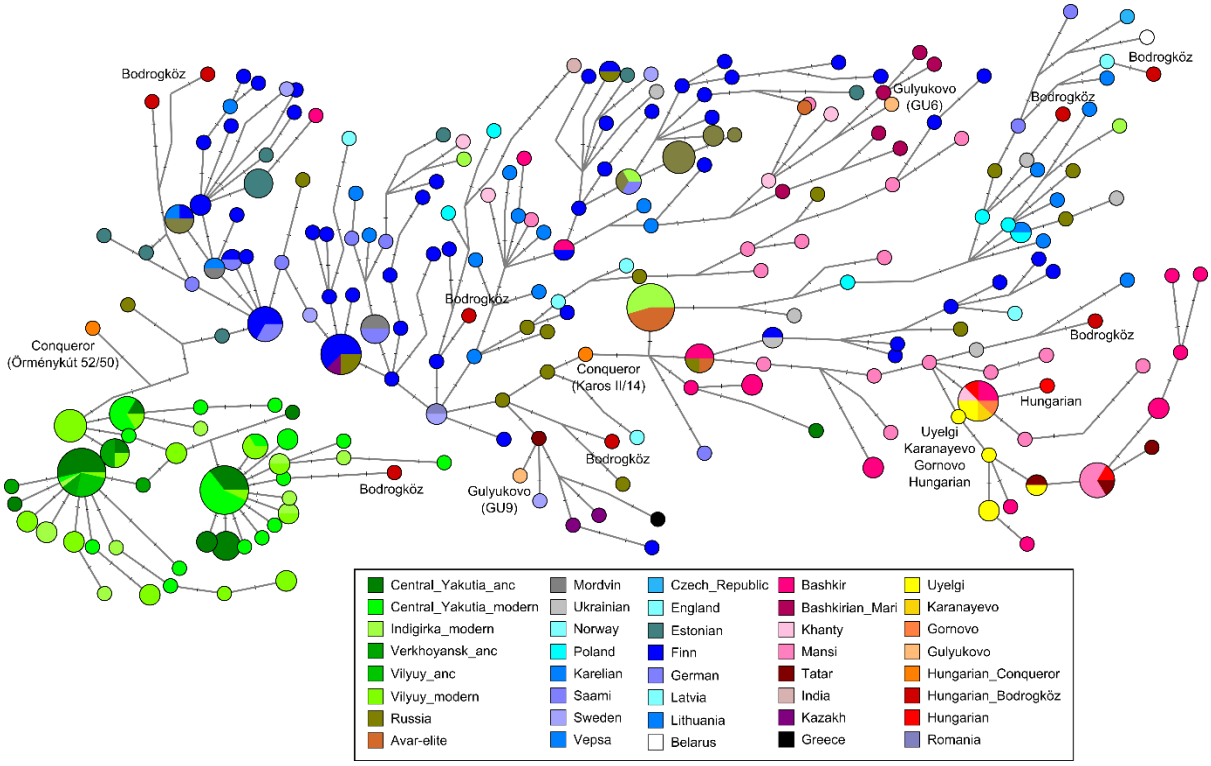

Figure 3: Median-joining network of N1a1 Y-chromosomal haplogroup based on 17 STRs. The studied Y lineages from Uyelgi+Karanayevo (Uyelgi and Karanayevo sites) and Chiyalik (Gornovo) groups are identical or closely related to each other and Bashkirs from Perm, Strelibashevsy, Burzyansky and Western Orenburg regions, as well as Khantys, Southern and Northern Mansis, modern-day Hungarians, Tatars and Ukrainians. One sample from the Chiyalik group (Gulyukovo site, GU6) is on a subbranch, mainly composed of Bashkirian Maris. The other sample from here (GU9) is on another subbranch, which contains Tatars and Russians as well. For information on the groups and STR markers see Supplementary Material, Table S10.

Mitochondrial haplogroup frequency based analyses

On the HG frequency based Principal Component Analysis (PCA) plot, the examined groups are in an intermediate position between populations of Eastern and Western Eurasia, which reflects that the Volga–Ural region served as a contact zone (Fig. 4, Supplementary Material, Fig. S36, Table S3). Because of their similar macro-haplogroup composition (Fig. 2), Uyelgi+Karanayevo, Chiyalik, Tankeevka, Bolshie Tigani and Novinki groups form one unit. Due to the lack of data for the region we sampled, this unit is mapped in a sparsely covered area of the plot most closely to modern Turkmens,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Uzbeks, Khantys, Mansis, and Iron and Bronze Age groups of Central Asia, pointing to a previously unexplored genetic cluster between East and West. The Ward type clustering supports the HG level relationship among the groups we studied. Most of them fall in the same cluster, on the main branch comprising steppe groups (Supplementary Material, FigS37). These results demonstrate that the Iron and Bronze Age groups in the area determined the genetic composition of the territory under investigation, but since they are chronologically far from the examined medieval groups, we cannot infer direct succession. The ancient proto-Ob-Ugric group is plotted next to the North and Central Asian unit, in the vicinity of the representatives of Iron Age and Bronze Age cultures in Russia (Okunevo, Krotovo, Central Asian Late Iron Age cultures), as well as modern North Asian groups (Nganasan, Even, Evenk), which affinities are also confirmed by the Ward analysis (Supplementary Material, Fig. S37).

The conqueror Hungarians (KL-IV, V and VI) are markedly different from one another. Group KL-IV contains more typical eastern and less characteristic western mitochondrial lines (Fig. 2). For this reason it is the most closely related to the studied VUR's groups. Presumably, this cemetery group was used by the first and second generations of the conquerors pointing to no or limited admixture with the local inhabitants of the CB. In contrast, the KL-V-VI groups already show traces of genetic fusion; they are located on PCA plots between the European and Bronze Age steppe groups (Fig. 4, Supplementary Material, Fig. S36). Cemeteries belonging to the KL-V were used in the 10th century, where an early stage of genetic assimilation can be observed on PCA. Assimilation is more pronounced in KL-VI cemeteries, which were used up to the 11-12th centuries (37) and this group is clustered with European populations in the Ward analysis as well (Supplementary Material, Fig. S37).

The role of the representatives of the Iron Age Sargat culture in the development of the early Ugric people has already been discussed in genetic publications (19) (43). This group from present-day Russia (RUS-Sargat) was close to Group KL-IV in our analyses, however it should be noted that the Sargat group is a small sampleset (n=18).

In each analysis, modern Hungarians belong to the European populations, which demonstrates their admixture with Europeans since the time of the Hungarian Conquest (21).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

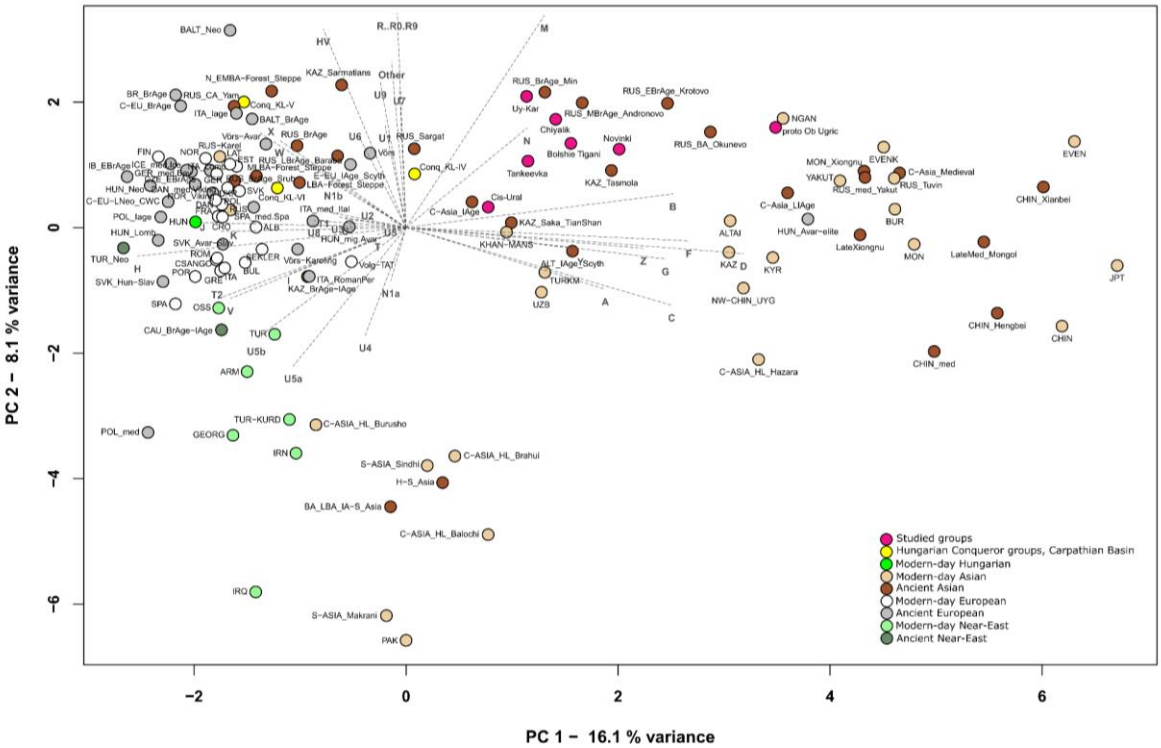

Figure 4: PCA plot based on mitochondrial haplogroup frequencies of ancient and present-day Eurasian and Near-Eastern populations. The variances presented on the first two components: PC1:16,1%; PC2: 8,1%. For abbreviations and information see Supplementary Material, Table S3. PC3 with a variance of 6% is presented on Supplementary Material, Figure S36.

Mitogenome sequence based analyses

The genetic distances (F_{ST}) are not significant between most of the studied Volga-Ural groups (Fig. 5, Supplementary Material, Table S4) except between the proto-Ob-Ugric and Tankeevka groups. The Uyelgi+Karanayevo group however differs significantly from the others with the exception of the proto-Ob-Ugric and Chiyalik groups. The reason for their outlying characteristics can be the numerous intra- and intersite genetic connections, which also indicate a close relationship between the two communities (Supplementary Material, Fig. S42).

Based on the results of the Mantel Test, the diversity of the examined VUR groups was not a cause of spatial distance among them, but rather large-scale mobility that influenced their maternal makeup (Supplementary Material Chapter B, 1.6., Table S6).

The three conqueror groups are related to the VUR groups in different ways (Fig. 5). The genetic distance of KL-IV is not significant from Cis-Ural, Bolshie Tigani, Tankeevka, and Chiyalik groups. In the case of the KL-V, this can only be said of the Cis-Ural and Bolshie Tigani groups. KL-VI is fundamentally different from all the other groups, except from the Tankeevka. This suggests that the population using the cemeteries classified in this group had already been mixed with the local populace of the CB, in line with the results of the haplogroup-based analyses. The distance between KL-IV and KL-V is not significant, which is also in agreement with the similarity of the materials found in these cemeteries (37). Based on progress over time, this may indicate relatively fast assimilation between the locals and the newcomers.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

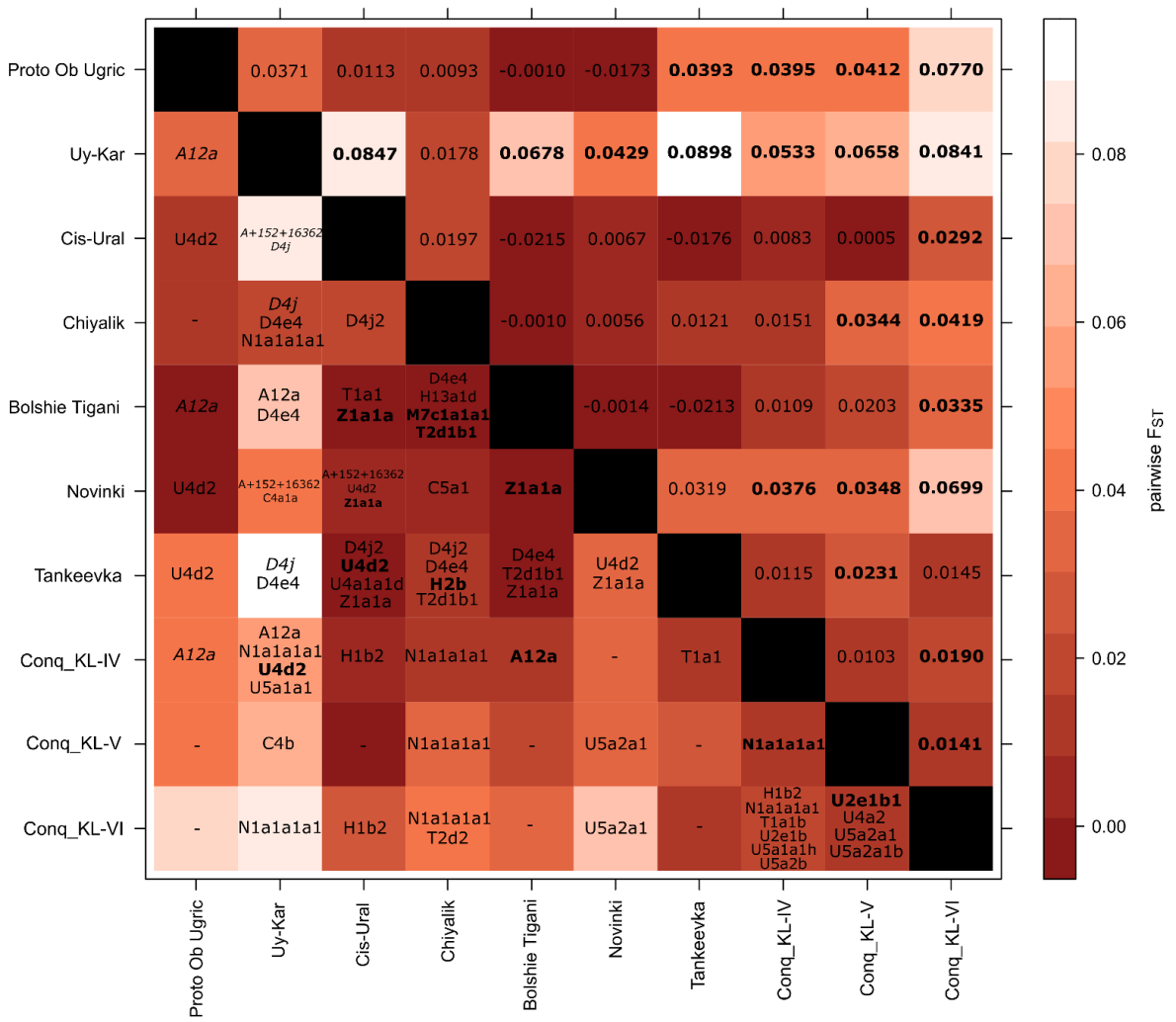

Figure 5: Shared mtDNA subhaplogroups between the studied groups (under the diagonal) and results of F_{ST} analysis (over the diagonal). F_{ST} values in bold indicate significant genetic distance. In the case of phylogenetic relationships, the more distal but unambiguous relationships are indicated by italics, the normal letters mark close relationships and letters in bold show haplotype identity between detected mitogenomes.

We performed clustering on ancient and recent Eurasian groups based on linearized Slatkin F_{ST} (Supplementary Material, Fig. S38), where ancient and recent populations form shared branches. Uyelgi+Karanayevo, Proto-Ob-Ugric and Novinki groups are in one cluster, which contains Southern and Central Asian Iron Age, Bronze Age and recent period populations as well. Bolshie Tigani and the Chiyalik groups are also in Southern/Central Asian context but in a different cluster than the earlier groups. Cis-Ural, KL-IV and KL-V are close to each other on a steppe origin subcluster, where the KL-V is closest to the Sargat population despite the previously demonstrated initial assimilation signals of the conquerors. Tankeevka and KL-VI are between modern-day and ancient European groups. For AMOVA analysis (Analaysis of Molecular Variance, Supplementary Material, Table S5) we classified the studied 10 groups into three sets based on the clustering results: 1) Uyelgi+Karanayevo, proto-Ob-Ugric, Novinki; 2) Tankeevka, Cis-Ural, KL-IV, KL-V, KL-VI; 3) Bolshie Tigani, Chiyalik. The source of variance among sets is 4.06% and their difference is significant (F_{CT}= 0.04058, p= 0.00782+-0.00313), while the variance within sets is 0.83% and they are not significantly different (F_{SC}= 0.00869, p= 0.15836+-0.01353). These results are confident with results of earlier analysis and with MDS

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Multidimensional Scaling) plot based on Slatkin F_{ST} values as well (Supplementary Material, Fig. S39).

Mitogenome phylogenetic analyses

The median-joining (MJ) network (Fig. 6.) based on the mitogenome sequences informs us about the diverse connections of the Hungarians in the CB and the studied VUR groups, which largely clusters as branches (haplogroups) of the mitochondrial phylogenetic tree (www.phylotree.org, Supplementary Material, Table S8., Chapter B, 2.2). In addition, the analysis also shows close intra-site maternal relationships.

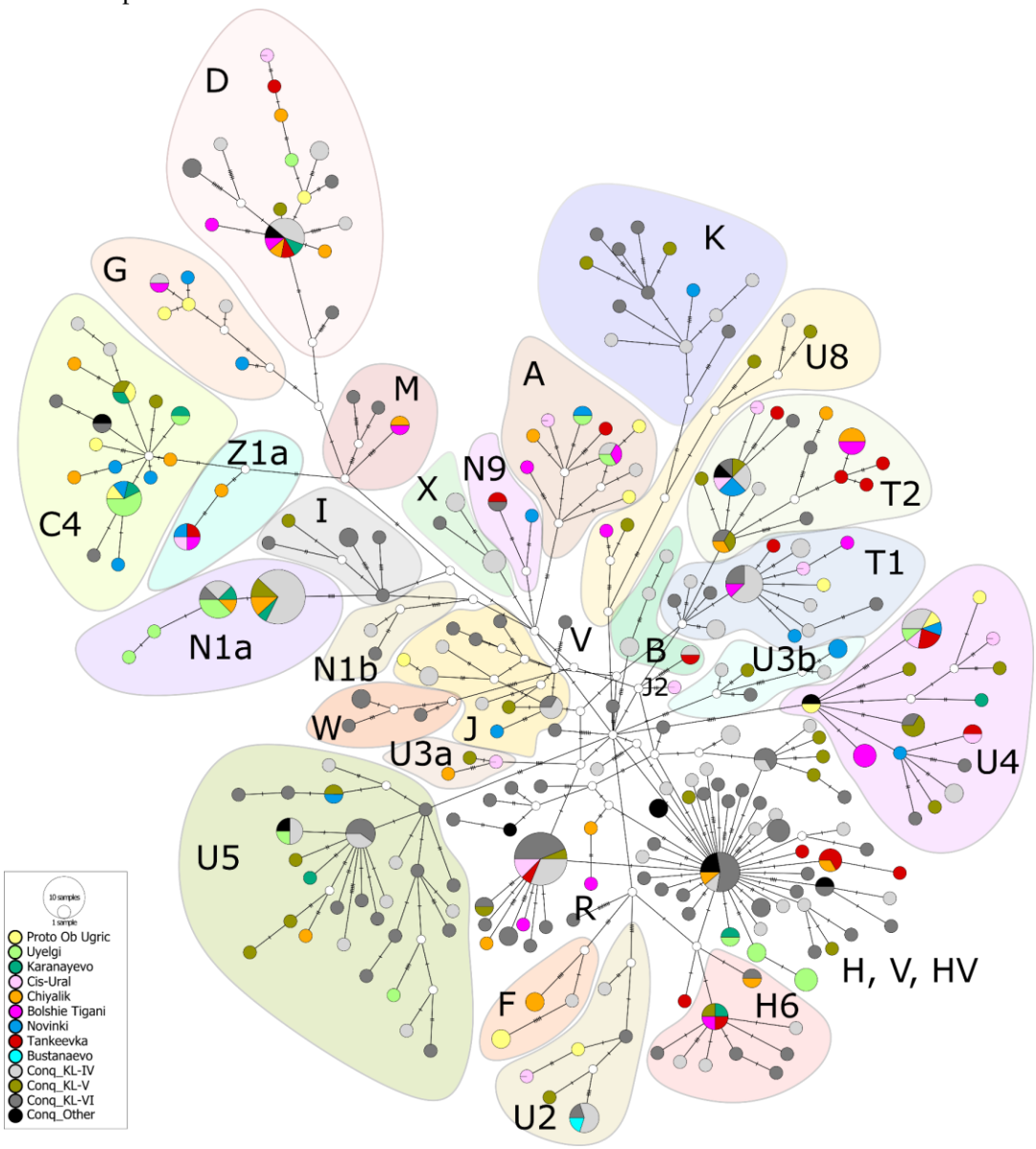

Figure 6: Median-joining network based on the maternal lines of the studied groups. This analysis shows a diverse system of relationships between our groups and the Conquest period Hungarians. Members of the studied groups are indicated by different colours, major mitochondrial haplogroups are labelled as colored clouds.

For a more detailed examination of maternal relationships, we made neighbour-joining (NJ) phylogenetic trees from those HGs separately that were detected in at least two groups associated with the Hungarians (including the Uyelgi, Cis-Ural, and KL-IV–VI groups, see Supplementary Material,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S8., Chapter B, 2.2). Inter-site haplotype (sequence) identities testify close relationships between the studied groups suggested also previously by archaeological finds (Supplementary Material, Fig. S42). We found the most numerous examples for this between the Bolshie Tigani and Chiyalik groups (within HGs M7c1a1a1, T2d1b1 (Supplementary Material, Fig. S50 and Fig. S54)), which is consistent with both HG frequency and F_{ST}-based analyses.

The phylogenetic trees of several HGs clearly demonstrate the existence of related maternal lineages among the groups under investigation, supporting their assumed connection (Fig. 5-6 and Supplementary Material, Chapter B, 2.2). Bolshie Tigani, Tankeevka and Chiyalik groups have the largest number of phylogenetic links. Furthermore, Cis-Ural and Tankeevka, as well as Chiyalik and Uyelgi-Karanayevo groups reveal many closely related lineages (Supplementary Material, Fig. S42). From the conquerors, KL-IV is connected with the largest number of maternal lines to the studied groups of the VUR (especially to the Uyelgi-Karanayevo) confirming the putative genetic relatedness behind the similarity of the archaeological finds. KL-V and KL-VI however are only loosely connected to the others, sharing no more than 1–2 lines.

The finding of many phylogenetic links between significantly different groups suggests that although the chronological difference and geographical distance, or possible admixtures with different groups transformed the overall genetic picture, common lineages were preserved at individual level in the descendants. This can be observed, for example, between Uyelgi-Karanayevo and KL-IV, between KL-IV and KL-VI, as well as between KL-V and KL-VI (Fig. 5-6.).

Majority of maternal lines can be traced back in Central Asia (south of the line connecting the Caucasus and present-day Kazakhstan; e.g. C4b, C4+152, U4a2) and the steppe areas (the forest-steppe and steppe grassland north of the former area through the Altai Mountains to Lake Baikal; e.g. C5c, C4a1a+195) according to phylogenetic analyses (Supplementary Material, Chapter B. 2.2). Several maternal lines of Siberian (e.g. A8a1, D4j4, T2d1b1, Z1a1a) and Middle Eastern (e.g. U3a, U3b) origins have also been detected, while some others point to the Far East (e.g. D4g1b, M7c). This diversity is expected on the basis of migrations going through Central Eurasia (18) (19) and is in line with the results of our population genetic analysis. Several maternal lineages represent a direct link between the groups from VUR and CB (e.g. A12a (Supplementary Material, Fig. S43A), N1a1a1a1a (Fig. 7 and Supplementary Material, Fig. S52), which proves that these groups most probably had a common source in (or near) the study region. The subhaplogroup N1a1a1a1a is particularly common among the Hungarian related groups (with a prevalence of 26% in Uyelgi+Karanayevo, 12% in Chiyalik group, 9,57% in KL-IV, and 5,26% in KL-V) and the structure of the phylogenetic tree suggests extended and almost exclusive maternal connections within and among these studied groups (and a representative of KL-VI). Based on the N1a phylogeny, the Kushnarenkovo and Chiyalik/Hungarian conqueror split can be dated to 600-750 AD, while Chiyalik and conqueror split date (9th century) coincides with the Hungarian conquest. This latter observation suggests rapid movement from the VU region to the CB that might have taken place within a generation.

The majority of the “oriental” elements previously identified in the conquerors could be detected in the Volga-Ural region, so they may as well have come directly from this region, even if they can also be traced back to more distant regions (such as Altai Mountains, Lake Baikal). Most of the analysed maternal lineages of the proto-Ob-Ugric group were of Siberian origin (unlike other studied groups and the conquerors), but lines of steppe and Central Asian origin were also detected in them.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

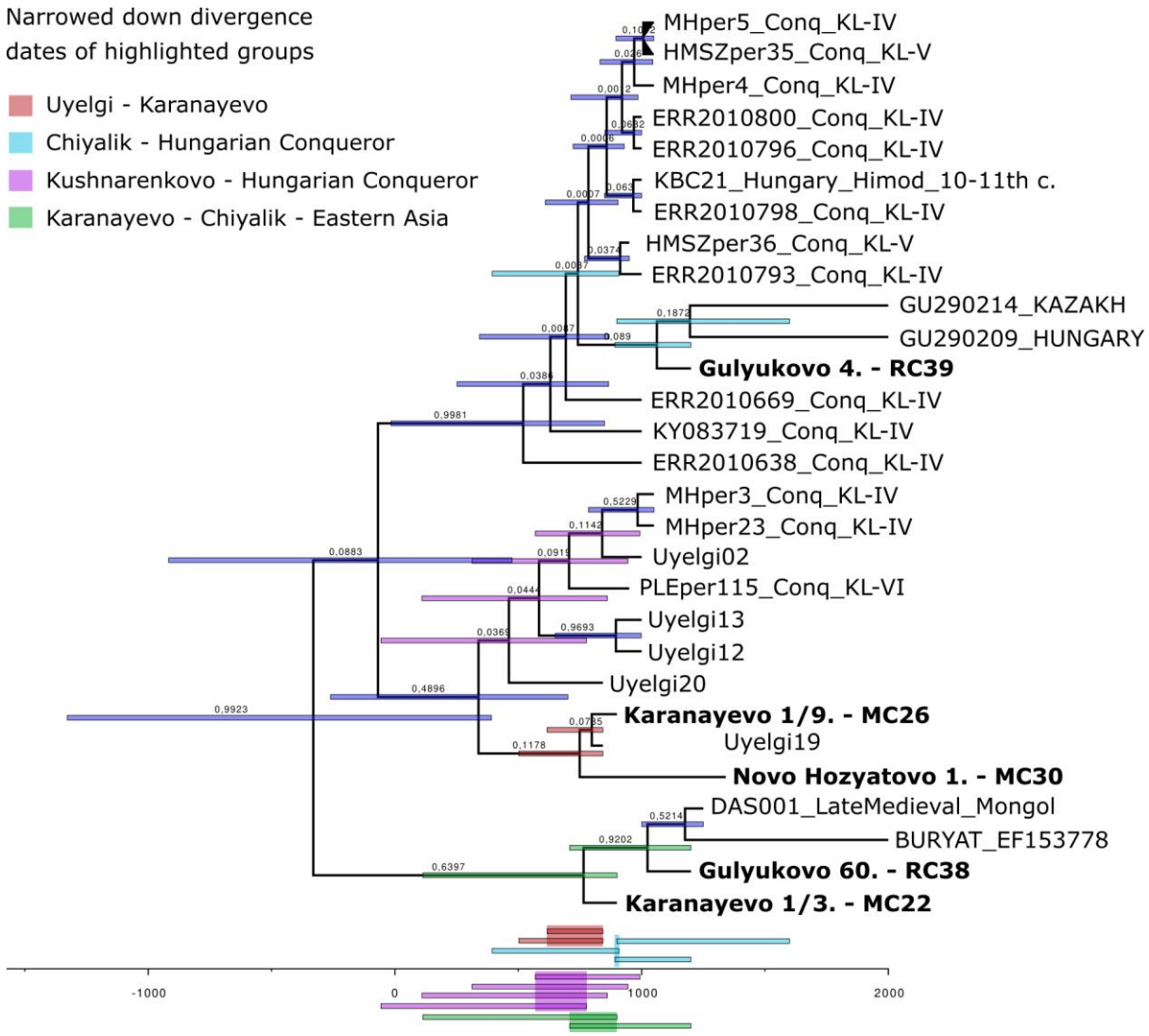

Figure 7: Mitochondrial haplogroup N1a1a1a1a phylogenetic tree with divergence dates, made with BEAST software. This tree mainly consists of samples linked to Hungarian prehistory including both database and our data (in bold). Divergence date estimations are correlating with our presumptions of group split times based on historical and archaeological data. Majority of the group divergence dates are 600-800 AD pointing to rapid population movements in this period. For the extended figure, see Extended Figure 3.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusions

This study reveals the first genetic data of the early medieval sites in Western-Siberian and Volga-Kama regions in addition to our previous publication from the Ural region (23). These studied regions are extremely important in the ethnogenesis of several Uralic and Turkic peoples, among them also the Hungarians, however, only a few of these population have been genetically studied so far.

Our analyses indicate no or little biological connection between the ancestors of Hungarians and proto-Ob-Ugric groups despite their close geographical proximity for 1500–2000 years after the date of their split that was estimated by linguistic models and chronology (9). We identified only a few uniparental links between them, but our results also show that there is a need for further studies of the proto-Ob-Ugric peoples as well as nomadic peoples that arrived in Western-Siberia from the south and east in several waves, who provided local ancestral components of later Western Siberian populations.

On the other hand, our results support the proposed intensive relationship between the eastern and western Uralic communities in the 9th–11th centuries (Uyelgi and Karanayevo sites) in the late Kushnarenkovo culture (44) (45) (46). We detected closely related and identical sequences and haplotypes both in the case of the maternal and paternal lineages. Furthermore the estimated divergence time of the mitochondrial haplogroup N1a1a1a1a also supports their close connections.

The western sites of the Uralic Kushnarenkovo and Karayakupovo cultures located along the Rivers Volga and Kama (Bolshie Tigani and Tankeevka) show individual genetic links to each other (D4e4, T2d1b1, Z1a1a), although population genetics pull Tankeevka closer to the European groups. According to F_{ST} analysis, both groups are similar to Volga-Tatars, who live today in the studied region, suggesting that the base population of the VU had still it's imprint in present-day groups of the region.

Together with historical data, our results clearly connect the population of Bolshie Tigani cemetery (9th–10th centuries) to the representatives of the sites of the later Chiyalik culture (11th–14th centuries), as we detected numerous (identical) maternal lineages between them (D4e4, H13a1d, M7c1a1a1, T2d1b1). Based on the population genetic analyses, the two groups had elements of common origin, which suggest at least partial continuity of the population, but the chronological gap also explains diverging affinities.

The burials of Novinki-type sites are archaeologically attributed to the representatives of Bulgar and/or Khazar heterogeneous (presumably border guarding military) groups. The genetic links of this group with Central and Inner Asia are in line with the historical and archaeological facts that the Khazars came from the territory of the Western Turkic Khaganate (47). Based on our results and historical and archaeological data (48), it is plausible that the members of this community came into indirect or even direct contact with the Hungarians or other Permian (Cis-Uralic) people (e.g. as mitochondrial HG Z1a indicates) living in their geographical neighbourhood.

We re-classified the burial grounds of the Hungarian Conquest period in the Carpathian Basin using archaeological cemetery typology (37) (38). Phylogenetic observations suggest that (at least a part of) the conquerors separated from the representatives of Kushnarenkovo culture 600-750 AD, while the Chiyalik group and conqueror split date was in the 9th century. In the CB, the new settlers and the local population only started admixing in the second half of the 10th century. Group KL-IV (10th AD) looks more like the population of the VUR sites associated with the Hungarians, both at the level of individuals and the community. A significant part of their uniparental lineages can be derived from the Volga and Ural regions and they show affinity to the Iron Age Sargat culture's population, which all suggest that they have only limited interaction with the locally found population. The cemeteries belonging to groups KL-V and KL-VI reflect increasing genetic absorption and the effect of the local population substrate, but some maternal lineages originating in the east also survived in these groups. Based on the numerous maternal links we detected among the populations of the VU region and the groups of conquerors, we conclude that men and women conquered the Carpathian Basin together.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the Volga-Ural region, we could identify most of the "eastern" genetic traits of the conquerors (28) (29) (30) (35) (36). Therefore, the historical, linguistic, and archaeological assumption that these population elements (i.e. a part of the conqueror Hungarians) may have originated directly or indirectly from this area has been supported by our results.

Our results confirm that the conquerors, and even their predecessors living in the region of the Volga and the Southern Urals, formed a composite, mixed population (9) (6) (23) (49), and their genetic makeup was influenced by the base population of the area.

The highlighted N1a1-M46 Y-chromosomal lineage shows genetic link between Kushnarenkovo and Chiyalik culture, conquerors and modern-day Hungarians as well as the VU region's present-day groups (Bashkir, Tatar, Khanty, Mansi). This lineage is another piece of evidence that (at least a part of) the Hungarians came from the Volga-Ural region, from the territory of the Kushnarenkovo and Karayakupovo cultures, and it also shows the shared genetic history of the conquerors and the recent populations of the VU. The shared (and also chronologically and geographically debated) history of Hungarians and Bashkirs can only be presented here from the aspect of the paternal lines, but it is also plausible on the mitochondrial level, because a part of the Chiyalik population was most probably assimilated into the Bashkirs during the Middle Ages.

In order to gain a full spectrum on the origins of the Hungarians, the genetic analyses of early Hungarian cemeteries discovered in the areas west of the Volga and the earliest Hungarians arriving in the CB (i.e. the actual conquerors) need to be examined in the future. The autosomal analyses with higher information resolution will help understanding the processes of population transformation in more detail.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and Methods

Presentation of the examined cemeteries and sites

The Bustanaevo cemetery at the western foothills of the Southern Urals (6th-7th century, early Kushnarenkovo culture) represents the first generations of the early Kushnarenkovo population (including the early Hungarians, according to most of the researchers) moving there from the Trans-Urals (50). We also studied the Karanayevo site (9th–11th centuries, late Kushnarenkovo culture) located at the western foothills of the Urals, where the burial customs and grave-goods are the closest analogues of the Trans-Uralic Uyelgi site.

In the cemeteries of Bolshie Tigani (9th–10th centuries) and Tankeevka (10th–11th centuries) (Kushnarenkovo/Karayakupovo culture mixed with early Bulgars) a Hungarian ethnic component could be identified on the basis of burial customs and grave-goods. They can be considered as early Hungarians who did not migrate to the areas west of the River Volga (51) (52) (53) (2) (54). The Chiyalik culture (Gulyukovo, Novo Hozyatovo, and Gornovo cemeteries, 11th–14th centuries) may have been later communities of the Hungarians who remained in the East, in an already Islamised environment. Due to the Muslim religion, fewer funeral offerings were placed in the graves than before, but based on their types and shapes, they can be compared with the artefacts characteristic of the Hungarians. The Novinki-type (8th–9th centuries) of sites (Novinki, Lebyazinka, Malaya Ryazan, Mullovka, Shilovka and Brusyany) in the Samara Bend of the Volga may be suitable for studying the former neighbours of the Hungarians and early Khazar–Hungarian relations. These sites yielded the archaeological heritage of the population (presumably consisting of artificially organised communities of Bulgar and Khazar origins) settled there to protect the most important river crossing on the eastern border of the Khaganate (47). The examined sites along the left bank of the Volga may provide important information about the migration of the Hungarians due to the archaeological finds discovered near the Urals (presumably belonging to the early Hungarians) and it is possible to compare them with their former neighbours.

The study of the proto-Ob-Ugric samples (6th–13th centuries; Ivanov Mis, Panovo, Ust-Tara, Vikulovo, and Barshov Gorodok sites) from Western Siberia was not carried out for linguistic reasons, that is for a common Uralic origin with the Hungarians. The reason was that the ancestors of the Hungarians lived in the immediate vicinity of these Ugric-speaking peoples for 1500–2000 years, at least to the mid-sixth century (10) (6) (55), but more likely to the beginning of the 9th century (2) (4). Archaeological evidence of this (e.g. ceramics of Taiga origin) can be found in many cemeteries in the forest-steppe region, even in the Uyelgi cemetery (56).

For more information about archaeological background and the studied sites see Supplementary Material, Chapter I.

Sample collection

We sampled bones and teeth of 112 early medieval individuals from modern day Russia. These findings came from 18 different burial places (dated to 6-14th centuries AD), which represent different cultures (Fig. 1, Table 1, Supplementary Material, Table S1.). In most cases we extracted DNA from the petrous bone (*pars petrosa ossis temporalis*). If this part of the skull was not preserved, we used teeth and long bone fragments. We radiocarbon dated 34 samples in Poznan and Debrecen ATOMKI laboratories (Supplementary Table S1), along with $d^{13}C$ and $d^{15}N$ measurements for 30 samples performed in Poznan Radiocarbon Laboratory. BP dates were calibrated using OxCal 4.4 and calibration curve IntCal20 (57) (58). We found substantial freshwater reservoir effects (59) in the case of the Western Siberian communities (having $d^{15}N$ >13 ‰ and $d^{13}C$>-20 ‰ values) along with much older ^{14}C dates than the archaeological dating (see Supplementary Material, Chapter A.9). Due to this uncertainty, we followed

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the original archaeological chronology in all cases and present radiocarbon dates only as supplementary information (Supplementary Material, Chapter A, Table S1 and Table S11).

Laboratory conditions

We worked in spatially separated pre- and post-PCR laboratories. The pre-PCR laboratory is a dedicated ancient DNA laboratory (Institute of Archaeogenomics, Research Centre for the Humanities, Eötvös Loránd Research Network). In this, every workflow was performed in separate rooms under sterile conditions following well-established ancient DNA workflow protocols, as described by Csáky et al. (60). We determined the mitochondrial and Y-chromosome haplotype of the laboratory staff and compared these data with the ancient bone samples' results.

Preparation of bone fragments and teeth

We bleached the surfaces of the bones and teeth and took photos of them, thereafter we irradiated them with UV-C light. After that we gained powder from the bone fragments and teeth. We used two different methods for this: long bone fragments' and teeth' surfaces were cleaned with sandblasting and mechanically ground into fine powder in a mixer mill. This method was also used for 53 petrous bones (61). In the case of the other 40 petrous bones, powder was gained by direct drilling under sterile conditions (62).

Ancient DNA extraction and DNA library preparation

The DNA extraction was performed according to the protocol of Dabney et al. (63) with minor changes (61) from 80-100 mg of bone powder. The success of the DNA extraction was verified by PCR reaction (61).

During DNA library preparation we worked according to the protocol of Rohland et al. (64) with small changes. Half-UDG treatment was used for most of the samples, except for 20 samples where no-UDG treatment was applied (Supplementary Material, Table S1). We used unique P5 and P7 internal barcoded adapter combinations for each sample. The DNA libraries were amplified (TwistAmp Basic – Twist DX Ltd), purified (AMPure XP beads (Agilent)), checked the concentration and fragment size (Qubit 2.0, Agilent 4200 TapeStation System (Agilent High Sensitivity D1000 ScreenTape Assay)).

Hybridization capture and Next Generation Sequencing

We used hybridisation capture method to selectively enrich the entire mitochondrial genome. Y-chromosome SNP capture (61) (65) (564 SNPs on the Y-chromosome) was also used for 22 individuals, who had been previously genetically defined as male (Supplementary Material, Table S2 and Table S10). The bait production and amplification method described by Csáky et al. (60). We provided the captured as well as raw libraries for the shallow shotgun sequencing with unique iP7 and universal iP5 indexes for multiplex sequencing (66). We used Illumina MiSeq device for the NGS sequencing with Illumina MiSeq Reagent Kit v3 (150-cycles) and Illumina MiSeq Reagent Kit v2 (300-cycles).

Bioinformatic analysis

We used the same in-house pipeline for read processing as described in Csáky et al. (23) with minor changes (D. Gerber et al. 2022, manuscript in preparation), for results see Supplementary Material, Table S2. The bam and FastQ files were uploaded to ENA (https://www.ebi.ac.uk/ena/browser/home) under the accession number: PRJEB49842. Haplogroup determination for mitochondrial DNA (mtDNA) was performed by Haplogrep 2 (67) on fasta files that were called with a custom R script suited for archaic DNA.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Y-chromosome SNP and STRexamination

The Y-chromosome SNP analysis was described by Csáky et al (23) and classification was performed according to ISOGG v15.73. Y-chromosome profiles were determined by combining the results of STR and SNP data by using both nevgen.org and yleaf v1 software (68).

We investigated the short tandem repeats (STRs) of the Y-chromosome using the AmpFlSTR® Yfiler® PCR Amplification Kit (Applied Biosystems) and the subsequent data analysis and haplotype determination was carried out with GeneMapper® ID Software v3.2.1 (Applied Biosystems) (60). We performed independently repeated reactions on samples where at least 4 STRs were determined.

The prediction of Y-chromosome haplogroup was made by nevgen.org (https://www.nevgen.org/). We accepted as a valid result where the haplogroup probability was at least 50% and we could determine at least 8 STR loci (for major haplogroups), or we determined less than 8 loci, but the probability was more than 80% or SNP analysis confirmed the haplogroup (Supplementary Material, Table S2).

Based on Y-chromosome STR data we performed Median-joining (MJ) network analysis with Network v10.1.0.0 and the results we visualised with Network-Publisher v2.1.2.5. (Supplementary Material, Table S9).

Population genetic analyses

We performed different population genetic analyses, in which we compared the studied populations to several other ancient and modern-day populations.

For the PCAs we used mtDNA haplogroup frequencies of 69 ancient and 63 modern-day populations (Supplementary Material, Table S3). The PCAs were made in R v4.0.0. with prcomp and the results were visualised in two-dimensional plots.

For Ward hierarchical clustering we applied the same population dataset as for PCAs. Based on mtDNA haplogroup frequencies we used the Euclidean distance measurement method. We displayed the results as a dendrogram in R v4.0.0 with the pvclust library.

We calculated population pairwise F_{ST} and linearized Slatkin F_{ST} (69) values based on whole mitochondrial genome sequences of 3863 modern-day and 1741 ancient individuals using Arlequin v3.5.2.2. with the following settings: we used Tamura & Nei substitution model (70) with 10,000 permutations and significance level of 0.05. We applied 0.3 as a gamma value. The modern-day individuals were classified into 40, the ancient individuals into 46 distinct groups (Supplementary Material, Table S4).

To assess the correlation between genetic and geographic distances for the studied groups (Table 1), we performed Mantel test (71) based on pairwise F_{ST} (Supplementary Material Table S6). In groups that include multiple sites, a geographic centre was given to determine the distance from other groups or sites. For this analysis we used Arlequin v3.5.2.2.

The linearised Slatkin F_{ST} values (Supplementary Material, Table S4) were used for clustering, which we calculated in Python using the seaborn clustermap function with parameters correlation distance metric and complete linkage method for calculating the clusters.

MDS calculation was made based on linearized Slatkin F_{ST} values and results of it were visualised in two-dimensional plots calculated on Euclidean distances implemented in vegan library of R v4.0.0. (Supplementary Material, Table S4).

The linearised Slatkin F_{ST} values were used for AMOVA analysis in Arlequin v3.5.2.2, in which we created groups from this study, from published Uyelgi, Cis-Ural and Hungarian conqueror groups from Carpathian Basin (KL-IV, KL-V, KL-VI) (Table 1, Supplementary Material, Chapter B, Table S4, Table S5).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phylogenetic analyses

We drew and visualised the median-joining (MJ) network of the mitochondrial genomes of our investigated groups with the PopArt program. The input file of the PopArt was made by DnaSP. For this analysis we used 424 sequences, which contained 495 variable sites and belonged to 268 haplotypes.

To analyse maternal relationships between the newly investigated sites we prepared neighbour-joining (NJ) phylogenetic trees from those haplogroups, that were detected at least two different sites or groups of sites (including Uyelgi, Cis-Ural and conqueror groups) (Table 1). We used the method described in Csáky et al. (23) with a highly extended dataset, i.e. by using all known and available sequences assigned to a certain haplogroup. Phylogenetic trees show only highlighted subbranches of interest, see Supplementary Material, Table S9 and Chapter B.2.2.).

We checked the variants in Haplogrep 2 (67) of those haplogroups that have been described in more than one individual within a site.

BEAST v1.10.4 (72) was used to estimate divergence dates of the maternal lineage N1a1a1a1a, which has high prevalence in the most studied groups and also has relatively high mutation rate, which makes it suitable for such analysis. The following options were used: HKY nucleotide model, base frequencies were estimated according to nucleotide diversity, site heterogeneity model was Gamma + invariant sites where Gamma categories were set to 4. Clock type was set to random local, and we assumed a constant size model tree. We ran four MCMC of 100 million chains, which we merged afterwards with a 10M burnin and median height estimates of Bayesian posterior tree distribution. For divergence dates we used 95% highest posterior density weights.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Funding

This work was supported by the following fundings and grants: Russian Foundation for Basic Research (18-59-23002, 19-59-23006, 20-49-720010); Basic Research Program, Russian Academy of Sciences (121041600045-8); Thematic Excellence Program, National Research, Development and Innovation Office (TKP2020-NKA-11); Árpád dynasty program (IV.2).

Acknowledgement

We thank László Kovács and Nadin Rohland for their professional support, Dániel Budai for the coordinates of the studied sites and Viktor Szinyei and Zsóka Varga for preparing Fig.1.

The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of project No. 18-59-23002 “The origins of the formation of the culture of ancient Hungarians. Archaeological paleoanthropological and paleogenetic aspect of the study of medieval monuments of the Southern Urals and Western Siberia”.

The research was funded by RFBR [Russian Foundation for Basic Research] No. 19-59-23006: “The problem of cultural transformations of Magyars on the way of Hungarian Conquest”.

The research was funded by RFBR [Russian Foundation for Basic Research] No. 20-49-720010. The work was partially performed according to the Basic Research Program RAS No. 121041600045-8.

The research project has been realized within the project framework entitled: Archaeology Research on the Contacts between Hungary and the East (Our Eastern Heritage, PPCU History and Archaeology Interdisciplinary Research Team; TKP2020-NKA-11), with the support of Thematic Excellence Program, National Research, Development and Innovation Office and also within the project framework entitled: Árpád-dynasty program (IV.2).

Conflict of Interest Statement

The authors declare no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Botalov, S. G. (2013) Nekotorie aspekti uralskoi madjarskoi problemy. In Botalov, S. G. (ed.), *II-i Mezhdunarodnii Madjarskii simposium: sbornik nauchnyh trudov*, Chelyabinsk, pp. 139–167.
2. Komar, A. (2018) История и археология древних мадьяр в эпоху миграции. — A korai magyarság vándorlásának történeti és régészeti emlékei . Eds.: Türk, A. – Budai, D. *Stud. ad Archaeol. Pazmaniensia 11. – Magy. Őstörténeti Témacsoport Kiadványok 5. Budapest.*
3. Zelenkov, A. S. (2019) O statuse psevdokushnarenkovskoi keramiki v arheologicheskom kontekste Zapadnoi Sibirii v rannego srednevekovya. *Ab Orig. Arheol. Sb.*, **10**, 24–43.
4. Matveeva, N. P., Gyóni, G. and Zelenkov, A. S. (2021) Problemy izucheniya proishozhdeniya madjar (po uralo-sibirskim materialam rannego srednevekovya). *Ross. Arheol.*, 147–166.
5. Mazhitov, N. A. (1977) Yuzhnii Ural v VII–XIV vv. Yuzhnii Ural v VII–XIV vv. **(1977)** .
6. Ivanov, V. A. (1999) Drevnie ugri-madjary ревние угры-мадьяры в Восточной Европе. *Ufa*, Ufa.
7. Kristó Gy (1980) Levedi törzsszövetségétől Szent István államáig. *Levedi törzsszövetségétől Szent István államáig*; Magvető Könyvkiadó, (1980) .
8. Türk, A. and Langó, P. (2020) A magyarság korai történetének régészeti emlékei a legfrissebb leletek fényében. *A magyarság korai történetének régészeti emlékei a legfrissebb leletek fényében*; (2020) .
9. Türk, A. (2021) A korai magyar történelem régészeti kutatásainak aktuális eredményei és azok lehetséges nyelvészeti vonatkozásai (Recent advances in archaeological research on early Hungarian history and their potential linguistic relevance). In Klima, L., Türk, A. (eds.), *Parallel stories Interdisciplinary Conference on Hungarian Prehistory, organized by the Institute for Archaeology, PPCU, Budapest – 11–13 November 2020*, pp. 163–204.
10. Fodor, I. (1982) In Search of New Homeland: The Prehistory of the Hungarian People and the Conquest. *In Search of New Homeland: The Prehistory of the Hungarian People and the Conquest*; Corvina, Budapest, (1982) .
11. Györffy, G. (1986) Julianus Barát és napkelet fölfedezése. *Julianus Barát és napkelet fölfedezése*; Szépirodalmi Könyvkiadó, Budapest, (1986) .
12. Hautala, R. (2016) Early Hungarian information on the beginning of the Western Campaign of Batu (1235-1242). *Acta Orient.*, **69**, 183–199.
13. Kazakov, Y. P. (1987) O proishozhdenii i kulturnoi prinadlezhnosti pamyatnikov s grebenchato-shnurovoi keramikoi. *Problemy srednevekovoi arheologii Urala i Povolzhya*, Ufa.
14. Ivanov, V. A. (1984) Putyami stepnih kochevii. Putyami stepnih kochevii **(1984)** .
15. Garustovich, G. N. (2015) The Chiyalik Archaeological Culture in the Middle Ages on the South Urals. *Ufimskii arheologicheski vestnik*, pp. 181–198.
16. Kazakov, Y. P. (2007) Volzhskie bolgari, ugri i finni v IX – XIV vv. *Problemy vzaimodeistviya*, Kazan.
17. Kvitnitskii, M. V., Türk, A., Telnov, N. P., et al. (2020) Two Hungarian Graves of the 9th Century in the Danube-Dnister Steppes. *Strat. plus*, **5**, 329–340.
18. Narasimhan, V. M., Patterson, N., Moorjani, P., et al. (2019) The formation of human populations in South and Central Asia. *Science (80-.).*, **365**, eaat7487.
19. Gnecchi-Ruscone, G. A., Khussainova, E., Kahbatkyzy, N., et al. (2021) Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Sci. Adv.*, **7**.
20. Jeong, C., Wang, K., Wilkin, S., et al. (2020) A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell*, **183**, 890-904.e29.
21. Tambets, K., Yunusbayev, B., Hudjashov, G., et al. (2018) Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biol.*, **19**, 1–20.
22. Bermisheva, M. ., Tambets, K., Villems, R., et al. (2002) Diversity of Mitochondrial DNA Haplogroups in Ethnic Populations of the Volga–Ural Region. *GENOMICS. Transcr. PROTEOMICS*, **36**, 990–1001.
23. Csáky, V., Gerber, D., Szeifert, B., et al. (2020) Early medieval genetic data from Ural region evaluated in the light of archaeological evidence of ancient Hungarians. *Sci. Rep.*, **10**.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. Batini, C., Hallast, P., Vågene, Å. J., et al. (2017) Population resequencing of European mitochondrial genomes highlights sex-bias in Bronze Age demographic expansions. *Sci. Rep.*, **7**, 1–8.
25. Malyarchuk, B., Derenko, M., Denisova, G., et al. (2018) Whole mitochondrial genome diversity in two Hungarian populations. *Mol. Genet. Genomics*, **0**, 0.
26. Tombácz, D., Maróti, Z., Kalmár, T., et al. (2019) Whole-exome sequencing data of suicide victims who had suffered from major depressive disorder. *Sci. Data*, **6**, 1–10.
27. Post, H., Németh, E., Klima, L., et al. (2019) Y-chromosomal connection between Hungarians and geographically distant populations of the Ural Mountain region and West Siberia. 1–10.
28. Neparáczki, E., Kocsy, K., Tóth, G. E., et al. (2017) Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing. *PLoS One*, **12**, 1–11.
29. Neparáczki, E., Maróti, Z., Kalmár, T., et al. (2018) Mitogenomic data indicate admixture components of Central-Inner Asian and Srubnaya origin in the conquering Hungarians. *PLoS One*, 208.
30. Maár, K., Varga, G. I. B., Kovács, B., et al. (2021) Maternal lineages from 10-11th century commoner cemeteries of the Carpathian Basin. *Genes (Basel).*, **12**.
31. Tömöry, G., Csányi, B., Bogácsi-Szabó, E., et al. (2007) Comparison of Maternal Lineage and Biogeographic Analyses of Ancient and Modern Hungarian Populations. *Am. J. Phys. Anthropol.*, **76**, 989–993.
32. Csősz, A., Szécsényi-Nagy, A., Csákyová, V., et al. (2016) Maternal Genetic Ancestry and Legacy of 10th Century AD Hungarians. *Sci. Rep.*, **6**, 33446.
33. Neparáczki, E., Juhász, Z., Pamjav, H., et al. (2016) Genetic structure of the early Hungarian conquerors inferred from mtDNA haplotypes and Y-chromosome haplogroups in a small cemetery. *Mol. Genet. Genomics*, **292**, 201–214.
34. Kivisild, T., Shen, P., Wall, D. P., et al. (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373–87.
35. Fóthi, E., Gonzalez, A., Fehér, T., et al. (2020) Genetic analysis of male Hungarian Conquerors: European and Asian paternal lineages of the conquering Hungarian tribes. *Archaeol. Anthropol. Sci.*, **12**.
36. Neparáczki, E., Maróti, Z., Kalmár, T., et al. (2019) Y-chromosome haplogroups from Hun, Avar and conquering Hungarian period nomadic people of the Carpathian Basin. *Sci. Rep.*, **9**.
37. Kovács, L. (2013) A Kárpát-medence honfoglalás és kora Árpád-kori szállási és falusi temetői. *A honfoglalás kor kutatásának legújabb eredményei : tanulmányok Kovács László 70. születésnapjára*, pp. 511–604.
38. Révész (2020) A 10–11. századi temet˝ok regionális jellemz˝oi a Keleti-Kárpátoktól a Dunáig. *A 10–11. századi temet˝ok regionális jellemz˝oi a Keleti-Kárpátoktól a Dunáig*; Szegedi Tudományegyetem Régészeti Tanszéke–Magyar Tudományos Akadémia Régészeti Intézete–Magyar Nemzeti Múzeum–Martin Opitz Kiadó: Budapest, Hungary; Debrecen, Hungary, 2020, (2020) .
39. Grugni, V., Raveane, A., Ongaro, L., et al. (2019) Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC Biol.*, **17**, 1–14.
40. Fehér, T., Németh, E., Vándor, A., et al. (2014) Y-SNP L1034: limited genetic link between Mansi and Hungarian-speaking populations. *Mol. Genet. Genomics*, **290**, 377–386.
41. Ilumäe, A. M., Reidla, M., Chukhryaeva, M., et al. (2016) Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *Am. J. Hum. Genet.*, **99**, 163–173.
42. Pamjav, H., Fóthi, Fehér, T., et al. (2017) A study of the Bodrogköz population in north-eastern Hungary by Y chromosomal haplotypes and haplogroups. *Mol. Genet. Genomics*, **292**, 883–894.
43. Bennett, C. C. and Kaestle, F. A. (2010) Investigation of ancient DNA from western siberia and the sargat culture. *Hum. Biol.*, **82**, 143–156.
44. Botalov, S. G. (2016) Botalov, S. G.: Historical-cultural horizons of the early Iron and Middle Ages in the Trans- Urals forest-steppe. Some aspects of a Magyar problem. In: Археология Южного Урала. Лес, лесостепь. Ранний железный век и средневековье (проблемы культурогенеза). In Botalov, S. G. (ed.), *Археология Южного Урала. Лес, лесостепь.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ранний железный век и средневековье (проблемы культурогенеза), Chelyabinsk, pp. 552–553.

45. Sungatov, F. A. (2016) Karanaevskii kurganii mogilnik. in Archaeology of the Southern Urals. Forest, forest-steppe (problems of cultural genesis) . Eds. Botalov, S. G. et al. Chelyabinsk 2016, 409 – 426. In Botalov, S. G. (ed.), *Archaeology of the Southern Urals. Forest, forest-steppe (problems of cultural genesis)*, Chelyabinsk, pp. 409–426.
46. Botalov, S. G., Lukinih, A. A. and Tideman, Y. V. (2011) Pogrebalnii komplex mogilnika Uyelgi – novii srednevekovii pamyatnik v Yuzhnom Urale. *Chelyabinskii Gumanit. nauchnii zhurnal*, **2**, 104–114.
47. Róna-Tas, A. (2007) The Kazars and the Magyars. In Golden, P. B., Ben-Shammai, H., Róna-Tas, A. (eds.), *The World of the Khazars: New Perspectives : New Perspectives. Selected Papers from the Jerusalem 1999 International Khazar Colloquium*, Brill Academic Publishers, Leiden-Boston, pp. 269–278.
48. Stashenkov, D. A. (2020) Drevnie vengri i ih okruzhenie v Samarskom Povolzhe. *Katalog arheologicheskii kollektsii*, Samara, p. 2020.
49. Fodor, I. (2015) Венгры: древняя история и обретение Родины. Пермь. Венгры: древняя история и обретение Родины. Пермь **(2015)** .
50. Kolonskih, A. G. (2020) Nauchnii otchet ob itogah arheologocheskih raskopok Bustanaevskogo kurgannogo mogilnika na territorii Burayevskogo rayona Respubliki Bashkortostan v 2019. godu. Ufa 2020. Ufa.
51. Halikova, E. A. (1976) Ранневенгерские памятники Нижнего Прикамья и Приуралья. *Советская археология 1976:3, 141–156.*
52. Khalikova, E. A. and Kazakov, E. P. (1977) Le cimitiere de Tankeevka. *Les anciens Hong. Ethn. a l'est. Budapest. Ed. Erdélyi, I.*
53. Halikova, E. A. (1972) Погребальный обряд Танкеевского могильника и его венгерские параллели. Eds.: Смирнов, А. П. и др. Москва 1972, 145– 160. *Проблемы археологии и древней истории угров*, 145–160.
54. Fodor, I. (1972) К вопросу о погребальном обряде древних венгров. Eds.: Смирнов, А. П. и др. Москва 1972,. *Проблемы археологии и древней истории угров*, 176–188.
55. Belavin, A. M., Ivanov, V. A. and Krylasova, N. B. (2009) Ugry v Preduralya v drevnosti i srednie veka. Ugry v Preduralya v drevnosti i srednie veka **(2009)** .
56. Botalov, S. G. (2017) Pogrebalnii kompleks Uyelgi i nekotorie naplyudeniya na predmet ugorskogo i madyarskogo kulturgeneza. In Türk, A. (ed.), *Conference of Young Scholars on the Migration Period. November 4–6, 2014, Esztergom Volume 2*, Budapest-Esztergom, pp. 267–334.
57. Ramsey, C. B. (1995) Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon*, **37**, 425–430.
58. Reimer, P. J., Austin, W. E. N., Bard, E., et al. (2020) The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0-55 cal kBP). *Radiocarbon*, **62**, 725–757.
59. Philippsen, B. (2013) The freshwater reservoir effect in radiocarbon dating. *Herit. Sci.*, **1**, 1–19.
60. Csáky, V., Gerber, D., Koncz, I., et al. (2020) Genetic insights into the social organisation of the Avar period elite in the 7th century AD Carpathian Basin. *Sci. Rep.*, **10**.
61. Lipson, M., Szécsényi-Nagy, A., Mallick, S., et al. (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, 114488.
62. Sirak, K. A., Fernandes, D. M., Cheronet, O., et al. (2017) A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques*, **62**, 283–289.
63. Dabney, J., Knapp, M., Glocke, I., et al. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.*, **110**, 15758–15763.
64. Rohland, N., Harney, E., Mallick, S., et al. (2015) Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **370**, 20130624.
65. Haak, W., Lazaridis, I., Patterson, N., et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, **522**, 207–211.
66. Meyer, M. and Kircher, M. (2010) Illumina sequencing library preparation for highly

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.*, **5**.

67. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., et al. (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.*, **44**, W58–W63.
68. Ralf, A., Montiel González, D., Zhong, K., et al. (2018) Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Mol. Biol. Evol.*, **35**, 1291–1294.
69. Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
70. Tamura, K. and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*, **10**, 512–526.
71. Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.*, 209–220.
72. Suchard, M. A., Lemey, P., Baele, G., et al. (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.*, **4**, 1–5.

Tables

Group number	Sites	Country	Chronology (century AD)	Culture/Period	Abbreviation	Sample (N)	Reference
*	Bustanaevo	Russia	6-7th	Early Kushnarenkovo	Bustanaevo	1	This study
1.	Bajanovo, Szuhoj Log, Bartim, Brodi	Russia (Ural)	3-10th	Nevolino/Lomovatovo	Cis-Ural	14	(23)
2.	Uyelgi, Karanayevo	Russia (Ural)	9-11th	Late Kushnarenkovo	Uy+Kar	33	(23), This study
3.	Bolshie Tigani	Russia (Volga)	9-10th	Late Kushnarenkovo/Karayakupovo, early Volga-Kama Bulghar period	Bolshie Tigani	18	This study
4.	Tankeevka	Russia	10-11th		Tankeevka	22	This study
5.	Gulyukovo, Novo Hozyatovo, Gornovo	Russia (Volga)	11-14th	Chiyalik	Chiyalik	25	This study
6.	Ivanov Mis I, Panovo I, Ust-Tara, Vikulovo, Barsov Gorodok	Russia (Western-Siberia)	6-13th	Ust'-Ishim, Nizhneobskaya, Potchevash	proto Ob Ugric	16	This study
7.	Novinki, Lebyazhinka, Malaya Ryazan, Shilovka, Mullovka, Brusyany	Russia (Samara bend)	8-9th	Novinki-type sites	Novinki	19	This study
8.	Magyarhomorog-Kónyadomb, Szegvár-Szőlőkalja, Karos-I, Karos-II, Karos-III, Oroshaza-Gorbicstanya, Szabadkigyos-Palliget, Harta-Freifelt, Kenezlo-Fazekaszug	Hungary	10th	Hungarian Conquest period	Conq_KL-IV	115	(23), (28), (29), (30)
9.	Homokmégy-Székes, Nagytarcsa - Homokbánya	Hungary	10th	Hungarian Conquest period	Conq_KL-V	38	(30)
10.	M3 161. site, Ibrány Esbóhalom, Magyarhomorog-Kónyadomb, Püspökladány-Eperjesvölgy, Sárrétudvari-Hízóföld, Szegvár-Oromdűlő, Nyíregyháza-Oros Megapark	Hungary	10-12th	Hungarian Conquest period/Árpádian Age	Conq_KL-VI	130	(23), (28), (29), (30)

Table 1. Information about groups formed in this study. The sites were grouped according to archaeological and chronological aspects. We investigated only one individual from the Bustanaevo site that was not used in population genetic analysis due to its outlying chronology. For other information see Supplementary Material Chapter A, Table S1-S2 and Table S7.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.04.478947; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abbreviations

AMOVA	Analysis of Molecular Variance
CB	Carpathian Basin
HG	Haplogroup
KL-IV	Group-IV based on Kovács, 10th century small cemeteries of the camps
KL-V	Group-V based on Kovács, 10th century cemeteries of villages with a large number of burials
KL-VI	Group-VI based on Kovács, The cemeteries of villages opened in the 10th century and used until the 11th and 12th centuries
MDS	Multidimensional scaling
MJ	median-joining
NJ	neighbor-joining
PCA	Principal Component Analysis
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
VUR	Volga-Ural region

Author contribution

Designed the study: B.G.M, A.T., A.Sz-N.
Performed the experiments: B.Sz.
Analysed the data: B.Sz., D.G., V.Cs.,
Collected archaeological and anthropological material: D.A.S. A.A.H., A.G.S., I.R.G., E.V.V., N.P.M., A.S.Z., A.V.S., K.G.K., V.V. I., B.A.K., F.A.S., A.G.K., S.G.B, I.G.V., B.G. M, A.T.,
Wrote the paper:
B.Sz., D.G., P.L., B.E., A.T, A.Sz-N.