Mol Genet Genomics
DOI 10.1007/s00438-016-1267-z

ORIGINAL ARTICLE

Genetic structure of the early Hungarian conquerors inferred from mtDNA haplotypes and Y-chromosome haplogroups in a small cemetery

Endre Neparáczki[1] · Zoltán Juhász[2] · Horolma Pamjav[3] · Tibor Fehér[3] · Bernadett Csányi[4] · Albert Zink[5] · Frank Maixner[5] · György Pálfi[6] · Erika Molnár[6] · Ildikó Pap[7] · Ágnes Kustár[7] · László Révész[8] · István Raskó[9] · Tibor Török[1]

Received: 15 July 2016 / Accepted: 25 October 2016

Abstract We applied ancient DNA methods to shed light on the origin of ancient Hungarians and their relation to modern populations. Hungarians moved into the Carpathian Basin from the Eurasian Pontic steppes in the year 895 AD as a confederation of seven tribes, but their further origin remains obscure. Here, we present 17 mtDNA haplotypes and four Y-chromosome haplogroups, which portray the genetic composition of an entire small cemetery of the first generation Hungarians. Using novel algorithms to compare these mitochondrial DNA haplogroups with other ancient and modern Eurasian data, we revealed that a significant portion of the Hungarians probably originated from a long ago consolidated gene pool in Central Asia-South Siberia, which still persists in modern Hungarians. Another genetic layer of the early Hungarians was obtained during their westward migrations by admixing with various populations of European origin, and an important component of these was derived from the Caucasus region. Most of the modern populations, which are genetically closest relatives of ancient Hungarians, today speak non-Indo-European languages. Our results contribute to our understanding of the peopling of Europe by providing ancient DNA data from a still genetically poorly studied period of medieval human migrations.

Keywords Ancient DNA · Early Hungarian · Iterative rank correlation · Self-organizing cloud algorithm · MDS mapping

Introduction

Hungarian is the only non-Indo-European language in Central Europe; nevertheless, modern Hungarians are genetically very similar to the surrounding European populations (Lahermo et al. 2000; Lazaridis et al. 2014). According to prevailing academic theory, the language derives from

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-016-1267-z) contains supplementary material, which is available to authorized users.

✉ Tibor Török
torokt@bio.u-szeged.hu

[1] Department of Genetics, University of Szeged, Szeged, Hungary

[2] Hungarian Academy of Sciences, Centre for Energy Research, Budapest, Hungary

[3] DNA Laboratory, Network of Forensic Science Institutes, Ministry of Justice, Budapest, Hungary

[4] Department of Forensic Medicine, University of Szeged, Szeged, Hungary

[5] Institute for Mummies and the Iceman EURAC, Bolzano, Italy

[6] Department of Biological Anthropology, University of Szeged, Szeged, Hungary

[7] Department of Anthropology, Hungarian Natural History Museum Budapest, Budapest, Hungary

[8] Department of Archaeology, University of Szeged, Szeged, Hungary

[9] Institute of Genetics, Biological Research Centre, Szeged, Hungary

Published online: 01 November 2016

the Hungarian conquerors, which arrived from the Pontic steppes, and occupied the Carpathian Basin between 895 and 905 AD (Róna-Tas 1999). The genetic origin of the Hungarian conquerors, and their relation to present-day Hungarians and other modern and ancient populations is not yet clear. In a previous study (Tömöry et al. 2007), mtDNA haplogroups (Hg-s) were reported from 27 samples of the early Hungarians, and later from additional 41 samples (Bogácsi-Szabó et al. 2008), revealing diverse composition of haplogroups with significant Asian affinity. The authors concluded that the present-day Hungarians are not decedents of the conquerors, but these samples were selected from 15 archaeological sites, therefore, might not be representative due to sampling bias, and were dated to different time intervals, representing various conqueror generations. As the conquerors rapidly intermingled with local populations, genetic data from their first generation provide essential information about their ancestry.

The Hungarians arrived as an alliance of seven tribes, which associated shortly before the conquest. The arrival of the Hungarians is archaeologically well documented; their typically small cemeteries are distinguished by characteristic grave goods (see "Archaeological background"). These cemeteries were suddenly abandoned by the middle of the tenth century, probably due to monarchical reorganization of the society with the spread of Roman Christianity. From the early Hungarian period, cemeteries with the richest tomb artefacts were unearthed in North-Eastern Hungary, between the Tisza and Bodrog rivers, at the boundary of village Karos (Révész 1996). Three very similar well-dated cemeteries were excavated here from neighbouring sand dunes, which were all used by the arriving first generation Hungarians, from the last years of the nineth century to the middle of the tenth century. In this study, we report the maternal haplotypes of 17 out of 19 individuals, and the paternal haplogroups of four out of 14 men buried in the small Karos III. graveyard. We could also verify kinship relations within the cemetery by STR analysis. This study provides the first representative genetic data set, including paternal information from the first generation Hungarians. We compare our ancient Hungarian data set to that of other ancient and modern Eurasian populations, to reveal their possible origin and genetic relationships.

Materials and methods

Archaeological background

The three cemeteries of Karos-Eperjesszög (GPS: 48.327874, 21.719835) represent outstanding examples of the typical early conqueror cemeteries located around the upper Tisza river region, with the richest archaeological findings of the period. The three cemeteries with identical findings were probably used by contemporary neighbouring communities (Révész 1996). Cemetery no. I. was discovered in 1936 during plowing, and only 13 graves could be excavated, tillage destructed other 40–50 graves. The other two intact cemeteries (no. II. with 76 graves and no. III. with 19 graves) were excavated between 1986 and 1990 (Révész 1996). The cemeteries could be accurately dated with dirhams issued by Ahmed al-Mustain (862–866), Ismail ibn Ahmed (893–894, 898–899, and 905–906), Ahmed ibn Ismail (910–911) Tahir ibn Mohamed (906–907) Nasr ibn Ahmed (917–918 and 912–920), and coins issued by Ludwig IV. (900–911), king Berengar (888–915), and imperator Berengar I (915–924) to the earliest Hungarian period.

The graves were arranged in rows and groups (Fig. 1), had a typical west–east orientation, and contained a surprisingly high proportion (64%) of males, who died naturally, many at an old age, and were buried with typical warrior grave goods. In two of the cemeteries remarkably rich precious metal grave goods designated the tomb of the chief, but grave goods from other tombs also indicated that most deceased were distinguished members of their society. The unusually high proportion of males with weaponry and aristocratic goods implies that these communities might have been artificially organized military units, perhaps escort warriors of the ruler, with their servants and families, who were likely participants of the raids in Western Europe and Byzantium.

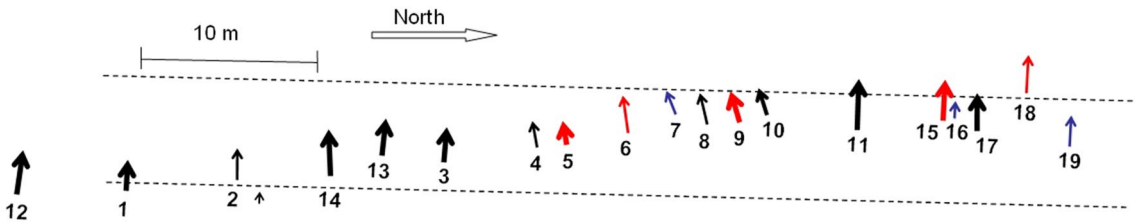

Fig. 1 Layout of the Karos-Eperjesszög cemetery no. III. *Arrows* indicate body orientation, individuals are identified with numbers. *Length of arrows* is proportional to grave size, *thick arrows* indicate elderly age, *blue colour* indicates children, and women are marked with *red*. *Dashed line* labels excavation segment (colour figure online)

Most of the man, some of the women and children were buried according to nomadic traditions, typically with horse cranium and leg bones, or adorned harness, saddle and stirrup, symbolizing horse (37 in cemetery no. II; 12 in cemetery no. III; and 31 estimated in cemetery no. I). Men wore precious metal jewels, penannular hair rings, band bracelet, and ball headed bezelled finger rings. Prestigious women were buried with silver and gold costume ornaments, braid ornament, ear pendants, and other jewels. Many of these objects are counterparts of archaeological findings from the East-European steppes and the Caucasus region. For example, similar sabretaches with metal plates were excavated around Kiev (Ukraine) and Martan-Ču (Russia, next to Georgia, Vinogradov 1983), and very similar caparisons are known from the Saltovo–Mayaki culture between the Don and the Dnieper Rivers (Pletneva 1981). Based on archaeological findings in the Karos III. cemetery, the following groups can be distinguished:

1. Leader (grave no. 11) with insignia of rank; waistbands with impressed precious metal belt-buckles, composite bow, mounted bow case, arrowheads, gilded-silver-mounted saber, horse cranium and leg bones, ornamented saddle and harness, metal-plated sabretache, stamped gold plate braid ornaments, boot mounts, golden coins, and jewellery.
2. Warriors (graves no. 2, 3, 4, 8, 10, 13, 14, and 17) with quiver and bow, arrowheads, battle axe, horse parts, parts of harness, stirrup, some with sabretache, and bronze belt-buckles.
3. Distinguished women (graves no. 5, 6, 9, 15, and 18) and children (graves no. 7, 16, and 19) with braid ornament, silver and gold costume ornaments, cast-bead row ear pendants, and other jewels.
4. Servants or commoners without findings (grave no. 1 and 12).

DNA extraction

Bone pulverization was performed, as described in (Benoit et al. 2013). Two or three DNA extractions were made from each sample with the silica suspension method of (Rohland and Hofreiter 2007), and DNA solubilisation was done overnight, in extraction buffer (EB) containing 0.45 M EDTA, 250 µg/ml Proteinase K, 1% Triton X-100, and 50 mM DTT.

Extraction no. 1 was made from 200 to 400 mg bone powder of femur or metatarsus bones with 4 ml EB, and 16 ml binding buffer (BB) containing 5 M GuSCN, 25 mM NaCl and 50 mM Tris, and 150 µl silica suspension.

Extraction no. 2 was made from 100 to 150 mg bone powder of tooth roots, with proportionally smaller volumes of the above buffers.

Extraction no. 3 was made from 400 to 600 mg powder of compact bones, but bone powder was first washed with 8 ml 0.5 M EDTA for 1 h at 48 °C, then predigested in 4 ml 0,5 M EDTA, 100 µg/ml Proteinase K for 20 min at 48 °C (Damgaard et al. 2015). Extraction was done in 3 ml of the same EB as previously, and then, 12 ml BB (5.2 M GuHCl, 100 mM NaOAc, 32% isopropanol, and 150 µl silica suspension) was added to the extract.

The pH of all binding buffers were adjusted between 4 and 6 with HCl, and after 3 h binding at room temperature, silica was pelleted, and washed twice with 80% ethanol. DNA was eluted in 200 µl TE buffer, and precipitated with equal volume of isopropanol in the presence of 20 µl 5 M NaOAc (pH 5.2) and 2 µl LPA carrier (Bartram et al. 2009). The pellet was washed twice with 70% ethanol, and finally, the DNA was redissolved in 50 or 100 µl TE.

Extraction no. 4 in Bolzano was made from 250 mg bone powder, in 5 ml EB as above, but without Triton X-100, and DTT, and 10 ml BB (4 M GuHCl, 32% isopropanol, 4% Tween-20, 100 mM NaOAc, and 100 µl silica suspension). Wash buffer was 125 mM NaCl, 10 mM Tris, 1 mM EDTA, and 50% ethanol.

mtDNA haplotyping

We PCR amplified and sequenced the 413 bp HVR-I mtDNA control region (np 15,997–16,409) with four overlapping primer pairs, as described in (Haak et al. 2008), and a 255 bp segment of the HVR-II region (np. 172–327) with a primer pair described in the same paper. Sequencing was done twice, at least from two extracts of the same sample. Sequences were submitted to NCBI GenBank under Accession No: KX519447–KX519463.

In addition to HVR sequencing, 22 Hg diagnostic coding region SNPs were determined by the GenoCoRe22 assay of (Haak et al. 2010) at least from two extracts Online Resource 1 (ESM_1). In all PCRs, GoTaq G2 Hot Start Polymerase (Promega) was used instead of AmpliTaq Gold and BSA (NEB) instead of RSA.

Y-chromosome haplogroup determination

To identify the basic Eurasian haplogroups, we used the GenoY25 SNP assay of Haak et al. (2010), with the only difference that in the multiplex PCR reaction, the concentration of the following primer pairs was modified to balance signal intensity: *M168, M2*, *M89,* and *M17* were used in 0.03 µM; M304 and *M9* in 0.04 µM concentrations, *M242, M45, M343*, and *M175* were used in 0.05 µM; and *M35* in 0.1 µM concentrations. At least two independent extracts were amplified from each sample Online Resource 2 (ESM_2).

Some of the missing or ambiguous positions could be implemented via single-plex PCRs, using optimized conditions for the given primer pairs. All single-plex PCR reactions were conducted in 25-µl volumes using 2 µl DNA extract, 2.5 mM $MgCl_2$, 6 µM of each primers, 0.25 mM of each dNTPs, 1 mg/ml BSA, and 1 U of GoTaq G2 Hot Start Polymerase with Flexi buffer. Primer annealing temperatures were the following: 62 °C for *M170, M126,* and *M304*, 57 °C for *S21*, and 55 °C for *M172*. Thermocycling conditions were 95 °C 2 min, followed by 50 cycles of 95 °C 20 s, annealing as above for 30 s, 72 °C 30 s, and final extension at 72 °C for 5 min. Non-template PCR controls were done for each reaction. PCR fragments were purified from agarose gels, and SBE reactions were done with the same conditions, as in the GenoY25 experiments (Haak et al. 2010), except that BSA was added (1 mg/ml).

STR analysis

To reveal possible kinship relations and confirm Y-chr Hg results of the GenoY25 assay, we also analysed autosomal and Y-chr short tandem repeats (STRs), albeit with limited success Online Resource 3 (ESM_3). Y-chr DNA was amplified in the Network of Forensic Science Institutes, Budapest, with the PowerPlex Y23 amplification kit, including 23 Y-STR loci, according to instructions of the manufacturer (Promega). Fragment sizes and allele designations were determined with Genetic Analysers 3130 and 3500 (Life technologies, Foster City, CA, USA) using the GeneMapper ID-X v.1.4 software. Y-chr Hg-s were identified with the Haplogroup Predictor Program of FTDNA.

Autosomal STR analysis was performed in Department of Forensic Medicine, University of Szeged with a GeneAmp PCR System 9700 thermal cycler (Life Technologies) using the PowerPlex ESX 17 System (Promega) according to the manufacturer's recommendations. The amplified fragments were analysed on ABI PRISM 310 Genetic Analyser (Life Technologies) capillary electrophoresis instrument with the GeneMapper ID v3.2 software (Applied Biosystems) Online Resource 3 (ESM_3).

Authentication criteria

Ancient DNA work was performed in the specialized ancient DNA (aDNA) facilities of the Department of Genetics, University of Szeged, Hungary with strict clean-room conditions, according to appropriate criteria (Knapp et al. 2012). The post-PCR work was done in a laboratory located in a separated remote side-wing of the building.

Throughout the experiments, the multistrategy DNA decontamination procedure of Champlot et al. (2010) was applied, namely, water was irradiated with 5 kGy gamma ray, enzymes, dNTP and primers were treated with hl-dsDNase and buffers were UV irradiated. Several DNA extraction and non-template PCR blank controls were used, and only samples with negative control results were maintained.

DNA isolation and HVR haplotyping of three samples (15, 16, and 19) were replicated in the laboratory of EURAC Bolzano, with identical results. Two of the Y-chr Hg-s revealed by GenoY25 SNPs could also be corroborated by STR analysis performed independently in the Network of Forensic Science Institutes, Budapest.

Hg-s and haplotypes of laboratory staff were determined, and were not identical with any of the samples.

The GenoCoRe and especially GenoY electropherograms often contained double peaks indicating contamination, nevertheless from repeated purifications and sequencing, multiple SNaPshot reactions, optimized single-plex PCR and STR data Hg-s could be determined unequivocally in most cases. The GenoCoRe22 results were concordant with Hg assignments based on HVR sequence.

We have identified a large number of Asian mtDNA haplotypes, which are very rare in modern-day Hungarian populations, and could not be derived from contamination.

Population genetic analyses

From the available literature, we assembled an Hg frequency database, considering 26 Hg-s, which contained 111 modern ($n = 20748$) and 35 ancient ($n = 1072$) Eurasian populations Online Resource 4 (ESM_4). In the ancient data set, we pooled populations with similar distributions, such as European Early with Middle Neolithic or Late-Neolithic with Early Bronze Age, but maintained the time resolution of samples from the Carpathian Basin. The ancient data set also contains two overlapping Hungarian conqueror populations published previously (Tömöry et al. 2007; Bogácsi-Szabó et al. 2008).

For comparing Hg distributions of ancient and modern populations, we applied the Self Organising Cloud (SOC) algorithm, a clustering method, based on automatic search for the local condensation points of the multidimensional point system defined by the 26-dimensional Hg distribution vectors, which has been successfully applied in the fields of genetics and ethnomusicology in previous works (Pamjav et al. 2012, 2013; Juhász et al. 2015). In SOC learning process, the mean Hg distributions of the resulting clusters are calculated by averaging the data vectors belonging to a common cluster, because the mean Hg distribution is a much more reliable approximation of the common features of the cluster than the individual data.

The "iterative rank correlation" (IRC) algorithm can give a hint, whether Hg distributions of populations reflect recent or ancient admixtures. If subsets of Hg-s propagate simultaneously from a common geographical source, it

will result in simultaneously varying frequencies in target populations during subsequent migrations. We have shown in a previous study that a search for such jointly propagating Hg-s can be successfully accomplished using the ICR algorithm (Juhász et al. 2016). In this algorithm, the rank (frequency) correlation of two Hg-s is calculated from the rank hierarchies of the populations based on correlative frequencies of the given Hg pairs, and the subset of the populations, where the rank correlation of two Hg-s becomes maximal is determined by an iterative learning process. This algorithm defines the subset of populations which produce the highest rank correlation for the given pair of Hg-s and calculates the correlation value. High correlation of Hg-s in aDNA data indicates that joint spreading from a common source happened before the period under study. Thus, instead of analysing the origin and migration pattern of a single Hg, we are searching for Hg pairs, triads, quartets, etc., spreading together from a probably common geographic source into different regions.

Phylogenetic study

To examine whether phylogenetic relations of the identified haplotypes within mtDNA haplogroups indicate phylogeographic relations, we constructed Median-Joining networks, using the Network 5.0.0.0 program (Bandelt et al. 1999). Network construction was based only on the nucleotide variations in the HVR-I region of published haplotypes, since from aDNA only HVR-I sequence was available. Here, we present only the Hg T and B networks, which revealed information about the possible relation of Karos individuals.

Results

mtDNA haplotype data

We have successfully amplified mtDNA from 17 samples, and determined the maternal haplotypes based on complete HVR-I, partial HVR-II sequences, and 22 mtDNA-coding region SNPs [see Table 1 and Online Resource 1 (ESM_1)]. Two samples (from graves 2 and 7) were eliminated from the data set due to bad DNA quality, resulting ambiguous classification. The 17 individuals belonged to 7 major Hg-s and 13 haplotypes.

The most frequent Hg was *B*, which together with Hg *A* indicate that about 30% of the Karos population is genetically connected to Central and East Asia. The majority of Hg-s (*H*, *U*, *T*, *J*, *X*) are of Eurasian origin; however, it is remarkable that two individuals belong to subhaplogroup *H6*, which may also indicate Asian connection. *H6* originated in the Near East and the Caucasus, and today is most frequent in Central and Inner Asia (21%), especially so in

Table 1 Summary of the mtDNA mutations and the defined haplotypes in the examined range; HVR-I: nt. 16050-16400, HVR-II: nt. 190-309, coding region 22 SNPs of the GenoCoRe22 assay

Sample (grave no.)	HVR-I mutations (position −16000)	HVR-II and coding region mutations	Haplotype	Haplogrep (%)
1	183C 189C 217C	263G 7028T 9 bp del 11719A 14766T	*B4*	100
3	362C	239C 263G	*H6*	100
4	069T 092C 126C 261T	228A 263G 295T 7028T 11719A 12612G 14766T	*J1c7*	100
5	183C 189C 217C	263G 7028T 9 bp del 11719A 14766T	*B4*	100
6	189C	263G 7028T 9 bp del 11719A 14766T	*B4′5*	100
8	051G 189C 362C	263G 7028T 11467G 11719A 14766T	*U2e*	95.89
9	051G	263G 7028T 11467G 11719A 14766T	*U2*	100
10	304C	263G	*H5*	100
11	189C 223T 278T	195C 257G 263G 6371T 7028T 11719A 12705T 14766T	*X2f*	100
12	183C 189C 223T 290T 319A	235G 263G 4248C 7028T 11719A 12705T 14766T	*A*	100
13	189C	263G 7028T 9 bp del 11719A 14766T	*B4′5*	100
14	126C 163G 186T 189C 294T	195G 263G 7028T 11719A 13368A 14766T	*T1a*	100
15	069T 126C 362C	263G 295T 7028T 11719A 12612G 14766T	*J*	98.15
16	256T 270T	263G 7028T 11467G 11719A 14766T	*U5a*	100
17	362C	239C 263G	*H6*	100
18	126C 163G 186T 189C 294T	214G 263G 7028T 11719A 13368A 14766T	*T1a10a*	100
19	126C 163G 186T 189C 294T	214G 263G 7028T 11719A 13368A 14766T	*T1a10a*	100
Researcher	rCRS	n. d.	*H2a2a1*	100

Altaians (35%) and appeared in Europe only in the Bronze Age (Roostalu et al. 2007).

The single *X2f* maternal haplotype of the chief (sample 11) is of particular interest, as this haplotype is most probably of south Caucasian origin, and still has the highest frequency in this region, but is rare in Eastern Europe, Central Asia, and is virtually absent in the Finno-Ugric and Turkic-speaking people of the Volga-Ural region (Reidla et al. 2003). The Hg *H5* of sample 10 has a similar distribution to *X2f*, and it is most frequent in the Caucasus, forming over 20% of hg *H* gene pool in some ethnic groups, and is absent from Volga–Uralic Finno-Ugrians and Central Asian populations (Roostalu et al. 2007). *H5* in Europe is most frequent in Slovaks and French, but it was also found in one LBKT sample from the Neolithic Carpathian Basin (Szécsényi-Nagy et al. 2015), so it did not necessarily arrive with the conquerors. The *U2* haplotype is found at low (0.5–2%) frequency throughout Eurasia, but in the Karos cemetery, we found two unrelated *U2* individuals suggesting that this Hg might had above average frequency in this group. The highest frequency of *U2* today is found among small ethnic groups of the Volga-Ural and North Caucasus regions, but *U2* and *U5* belong to the oldest European lineages which are present from the Mesolithic, and imply limited phylogeographic information. The same is true for the Hg-s *J* and *T*, which are present in Europe since the early Neolithic. The three individuals with Hg *T1a* indicate that this haplogroup might have been present with above average frequency among early Hungarians, and it is notable that Hg *T1a* today has an elevated frequency among Hungarian speaking Seklers, compared to surrounding populations (2.25%, Brandstatter et al. 2007).

Population genetic analyses

Many of the published aDNA haplotypes are inevitably inaccurate, owing to the PCR-based method, which obviously distort sequence-based population comparisons, although haplogroup determination is feasible. Furthermore, ancient Hg distributions are based on a rather small sample size, rendering similarity measurements with most methods rather uncertain. Therefore, we applied a more reliable approach, the SOC, and IRC algorithms for revealing genetic relations among populations (Juhász et al. 2015, 2016).

The logic behind IRC is illustrated by a concrete example in Fig. 2. The points in the coordinate system of Fig. 2a represent the 16-element "correlation subset" from the 35 ancient populations in our database, wherein frequencies of Hg-s *H* and *U5* strongly correlate. Along the *x*-axis the frequency of *H*, while along the *y* axis, frequency of *U5* decreases. The coordinates of the points are given by the frequency-dependent "ranks" of the corresponding population in the hierarchies of *H* and *U5*. Due to the relatively strong correlation of *H* and *U5*, the ranks of the populations are similar in both hierarchies, and therefore, the cloud of the points is situated around the diagonal axis. As the lower ranks refer to higher frequencies, the populations

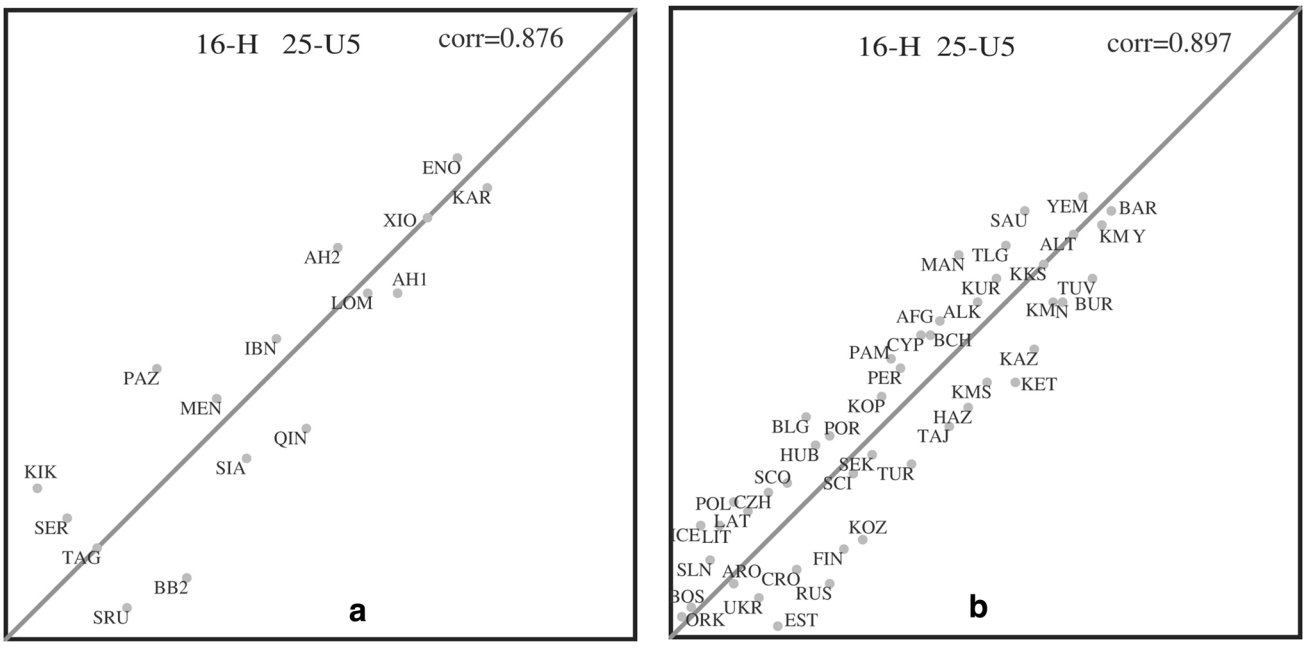

Fig. 2 Correlation diagrams of Hg-s *U* and *U5* in the ancient (**a**) and recent (**b**) databases (see text for explanation). Abbreviations of population names are given according to Online Resource 4 (ESM_4)

containing the largest components of Hg-s *H* and *U5* are situated in the vicinity of the pole. Applying the algorithm for a search in the recent part of our database (Fig. 2b), the strong correlation of Hg-s *H* and *U5* is also found within a much larger correlation subset containing 58 of the 111 recent populations.

The fact that the strong *H-U5* correlation can be detected both in ancient and recent data provided us the assumption that a significant part of recent populations may preserve ancient Hg distribution structures. Based on this assumption, we tested all Hg pairs with the IRC algorithm, whether they show similar connected frequencies, as above, in our ancient and modern database.

Accomplishing the IRC analysis first for the aDNA distributions, we obtained two well-separable disjunctive subsets of Hg-s having a strong positive rank correlation (covarying frequencies) among each other (Fig. 3a). Here, the "distances" of the Hg-s were determined, and the relationships were drawn in an MDS map trained by the resulting symmetric "distance" matrix of size of 26 × 26. We termed the larger subset containing Hg-s *I*, *H*, *J*, *T*, *U5*, and *K* "Western" and the smaller one containing Hg-s *C*, *D*, and *A* "Siberian" correlating Hg clusters (CHgC).

Accomplishing the same IRC learning process for the recent Hg distributions, we obtained four disjunctive CHgC-s (Fig. 3b), indicated by lines connecting pairs of Hg-s with a correlation beyond 0.75. It is meaningful that the ancient Western and Siberian CHgC-s were identified among the recent samples as well. It follows from this result that the migration processes determining the compositions of the ancient data also strongly determine the Hg distributions of a significant part of the recent populations in Eurasia. In other words, frequency correlation between members of the Western and Siberian CHgC-s was established before the age of the studied ancient populations, thousands of years ago, and subsequent population admixtures did not turn it over. As the further recent CHgC-s containing Hg-s *L*, *M**, *N**, and *R**, as well as *Y*, *U2* are not in contact with the two ancient CHgC-s, their analysis is beyond the scope of this study.

Next, we determined the subset of ancient populations containing the two correlating Hg clusters and represented their relations on MDS map, trained by the Euclidean distance matrix of their Hg distribution vectors (Fig. 4). According to the mathematical goal of the MDS algorithm, similar Hg distributions are mapped into neighbouring positions on the map [Online Resource 6 (ESM_6)], as it is illustrated by the lines connecting the Hg distributions having the smallest 26-dimensional Euclidean distances in Fig. 4.

The columns in Fig. 4a indicate the "fitness" of the Western CHgC in the ancient populations, and thus, the

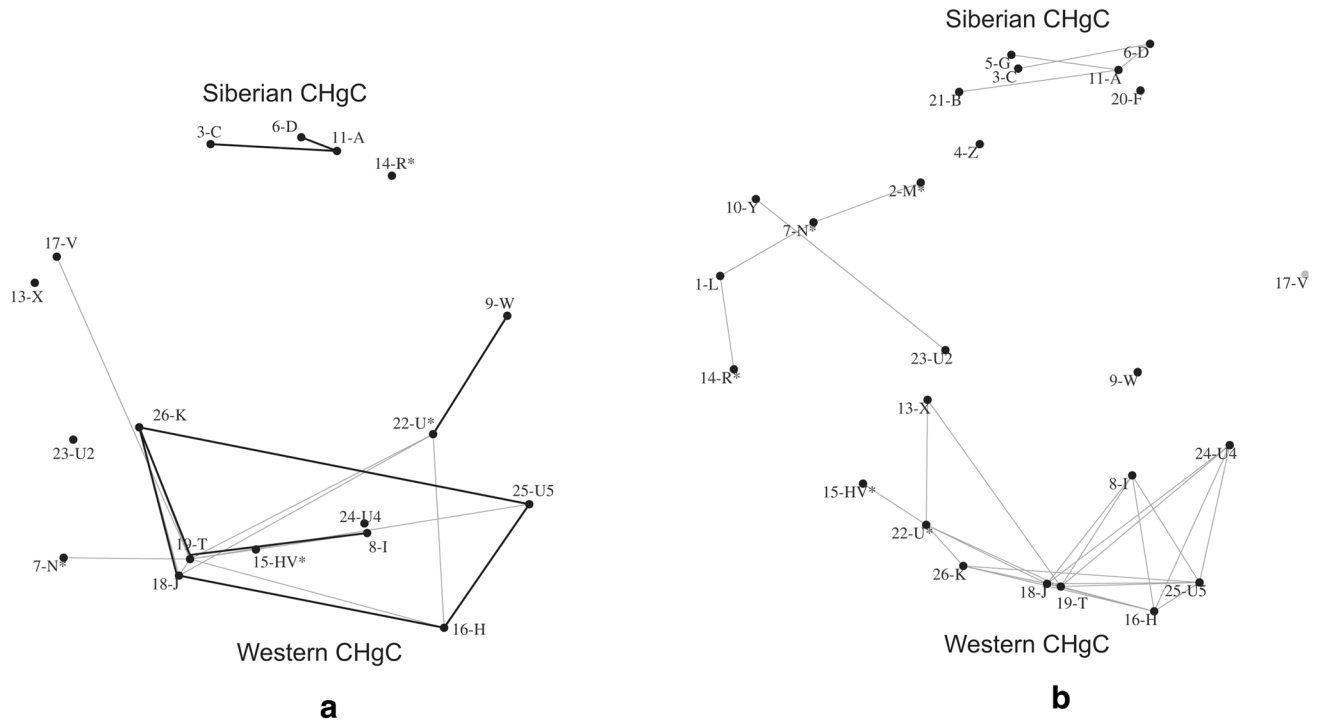

Fig. 3 MDS maps of the rank correlations of the 26 Hg-s in the ancient (**a**) and recent (**b**) databases. *Edges* connect pairs of Hg-s having the largest correlations. *Thick edges* connect Hg-s having a correlation beyond 0.77; *narrow edges* indicate correlations beyond 0.7 (see text for explanation)

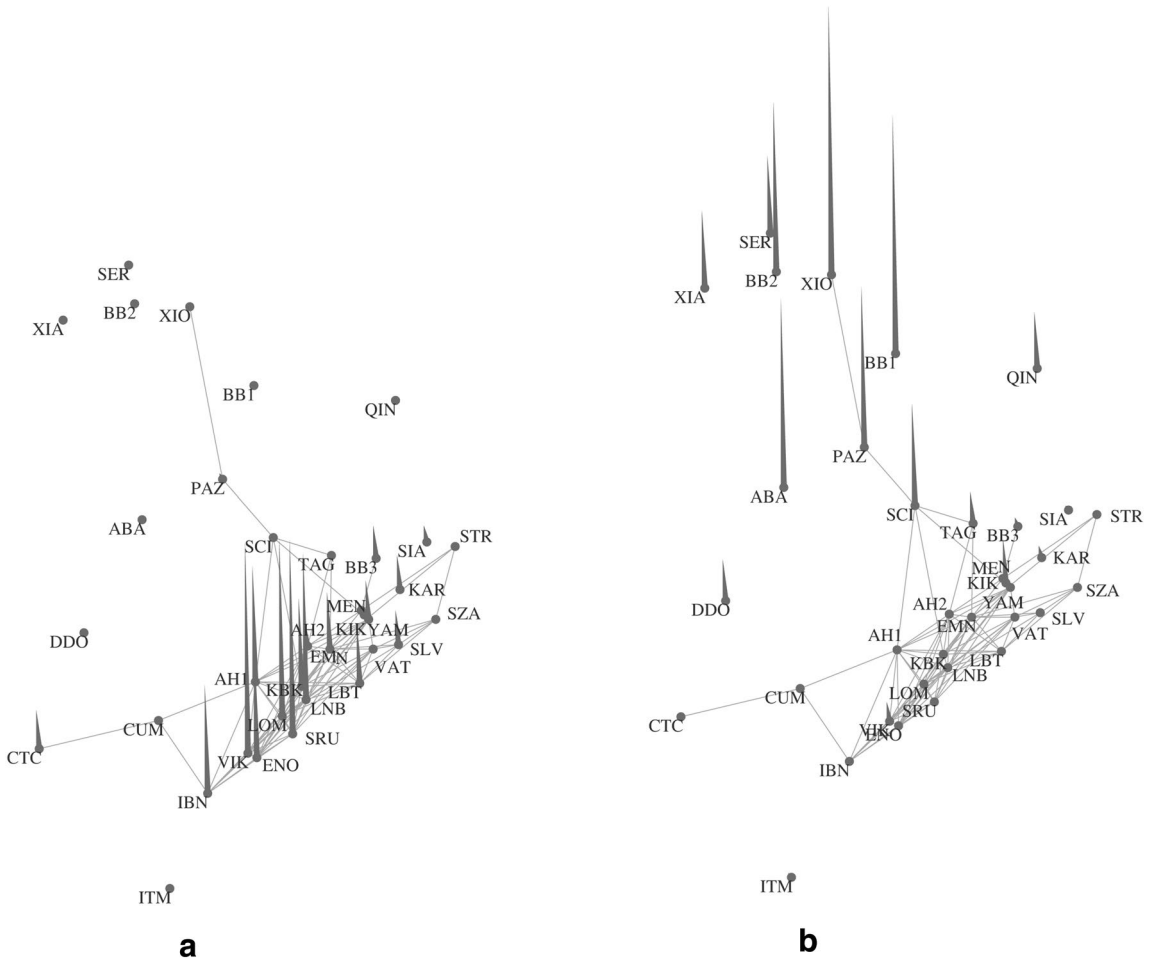

Fig. 4 MDS map of the ancient Hg distributions with the fitness values of the populations in the Western (**a**) and Siberian (**b**) ChgC-s. *Lines* connect populations with the smallest Euclidean distances (see text for explanation)

higher is the column the higher is the mean of the frequency of the "Western" CHgC in the corresponding population. For comparison, Fig. 4b illustrates the frequency of the "Siberian" CHgC, on the same MDS map. The fitness value (column length) was calculated as follows:

The normalised rank r_i of the ith population within the correlation subset of a pair of Hg-s is calculated as its rank divided by the highest rank value within the correlation subset. Thus, the normalised rank is a real number in the range between 0 and 1, where higher values mean lower frequency of the given Hg in the corresponding Hg distribution. It is obvious to define the "fitness" of the kth population of the given correlation subset as $I_k = E(1 - r_i)$, where E means the mean operator for all pairs of Hg-s having high correlation value within the CHgC and i is the index of the pairs of Hg-s having a strong correlation within the given CHgC.

Comparing the fitness values of the ancient populations within the Western (Fig. 4a) and the Siberian (Fig. 4b) CHgC-s, two groups can be distinguished containing mainly Western CHgC in the lower and Siberian CHgC in the upper part of the map. This result implies that two ancient "core populations", one with dominantly Western and the other with Siberian CHgC-s, probably strongly influenced the Hg distributions of other ancient populations by expansions and admixtures, and populations with high values are likely direct descendants of these "core populations".

Next, we used the IRC algorithm to similarly determine the complete subset of recent populations, which simultaneously contain members of the above two CHgC-s. The resulting subset of modern populations was probably significantly influenced by the two ancient "core populations". This subset was obtained as the union of all recent correlation subsets of Hg pairs belonging to either Western or Siberian CHgC. The size of this subset became 72 of the 111 recent populations, reducing our database to 35 ancient + 72 recent = 107 populations, which then resulted in clear and reproducible clusters. Based on the similarity of Hg distributions, the SOC algorithm identified

7 clusters of the 107 populations, where the average distance of the closest cluster centres became at least four times larger than the mean clustering error within clusters. Online Resource 5 (ESM_5) presents the characteristic distribution of haplogroup frequencies, which define the seven clusters, along with the list of populations belonging to each cluster.

The Karos population (KAR) belongs to cluster 2 (see ESM_5) which is characterized by the dominance of the Western CHgC, but the Siberian CHgC is also present in significant proportion. However, MDS map of cluster 2 (Fig. 5a) indicates that KAR is isolated within cluster 2, due to its relatively high Siberian components, and approaches cluster 3 which is a more balanced composite of Western and Siberian CHgC-s (ESM_5). In agreement with this, the closest and second closest mean Hg distributions to KAR are found in clusters 2 and 3; according to the smallest weighted Euclidean distances of Hg distributions (ESM_6), the closest relatives of KAR are AH2 (Hungarians 900 AD) GEO (Georgians), and UDM (Udmurts) all from cluster 2, and KUR (Kurdish) belonging to cluster 3. To illustrate the position of KAR to its closest relatives in clusters 2 and 3, we also present the combined MDS map of these clusters in Fig. 5b. Note that adding new components to an *n*-dimensional matrix, their MDS mapping in two dimensions will rearrange all components.

Phylogenetic study

Median-Joining network for Haplogroup *T* is shown in Fig. 6. The founder *T** haplotype was shared by 12 samples (2 Late Krotovo and 2 Fyodorovo samples from Baraba Steppe; 1 Khanty, 1 Altai Kazakh, 6 Bulgarians, and 1 Jordanian).

All three Karos conqueror *T* samples belonged to the *T1a** founder haplotype (red arrow on Fig. 6), which was shared by 43 other ancient and contemporary samples from the database; according to their origin these are 23 Bulgarians, 7 Hungarian speaking Seklers and Csángós, 2 Jordanians, 2 Mansi, 1 Khanty, 1 Altai Kazakh, 1 Han Chinese, 3 Yamnaya Kurgan aDNA, 1 Baraba Steppe Late Bronze Age aDNA, 1 Starcevo Neolithic aDNA, and 1 Transdanubian Linear Pottery aDNA. This finding confirms the genetic affinity between Hungarian conquerors and present-day Hungarians, as well as Bulgarians, who also originated from the same steppe region, and implies a phylogeographic link between the KAR population and ancient and modern populations in South Siberia.

The *T2b* founder haplotype contained 29 samples, including 13 contemporary Hungarian-speakers, 5 Bulgarians, 4 Neolithic samples from Starcevo, 4 Transdanubian Linear Pottery, 2 Cardium Pothery Neolithic samples, and 1 Eneolithic Kurgan sample from Ukraine. Four conqueror Hungarian aDNA published by (Tömöry et al. 2007)

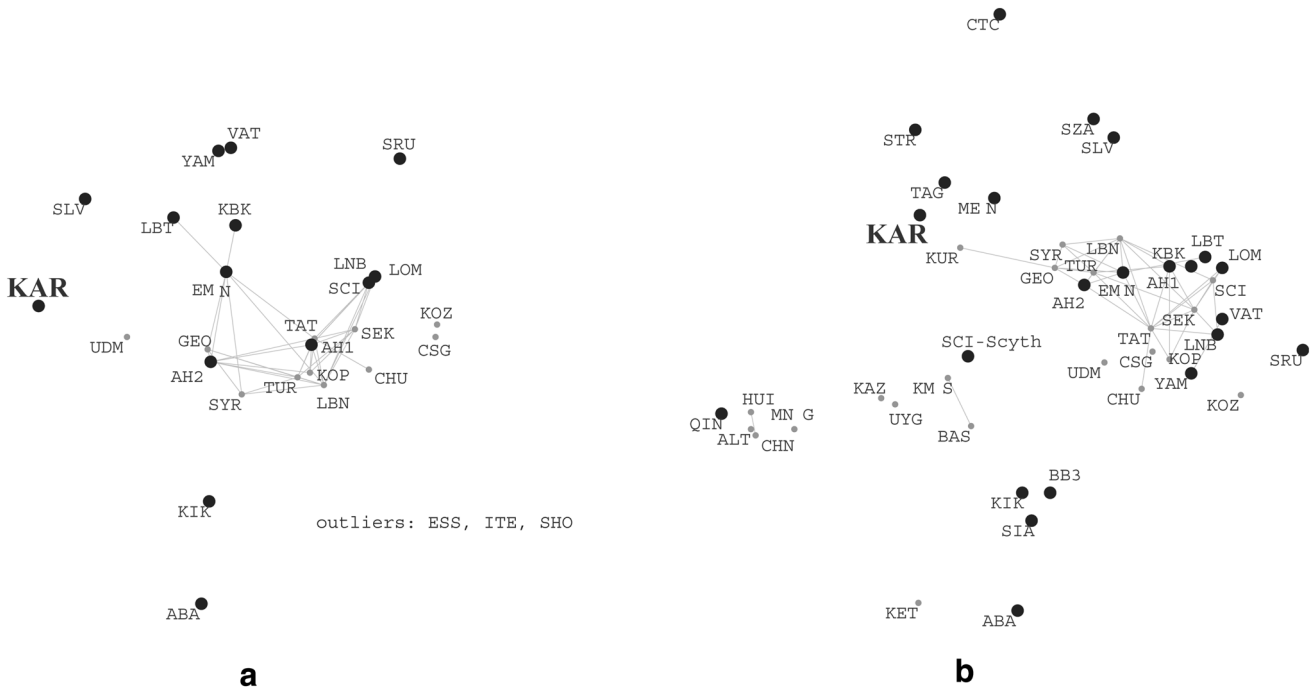

Fig. 5 **a** MDS map of cluster 2 Hg distributions. **b** MDS map of clusters 2 and 3 combined Hg distributions. *Lines* connect populations with the smallest distances. Ancient populations are labelled by *large black dots*, while recent populations are labelled by *small grey dots*

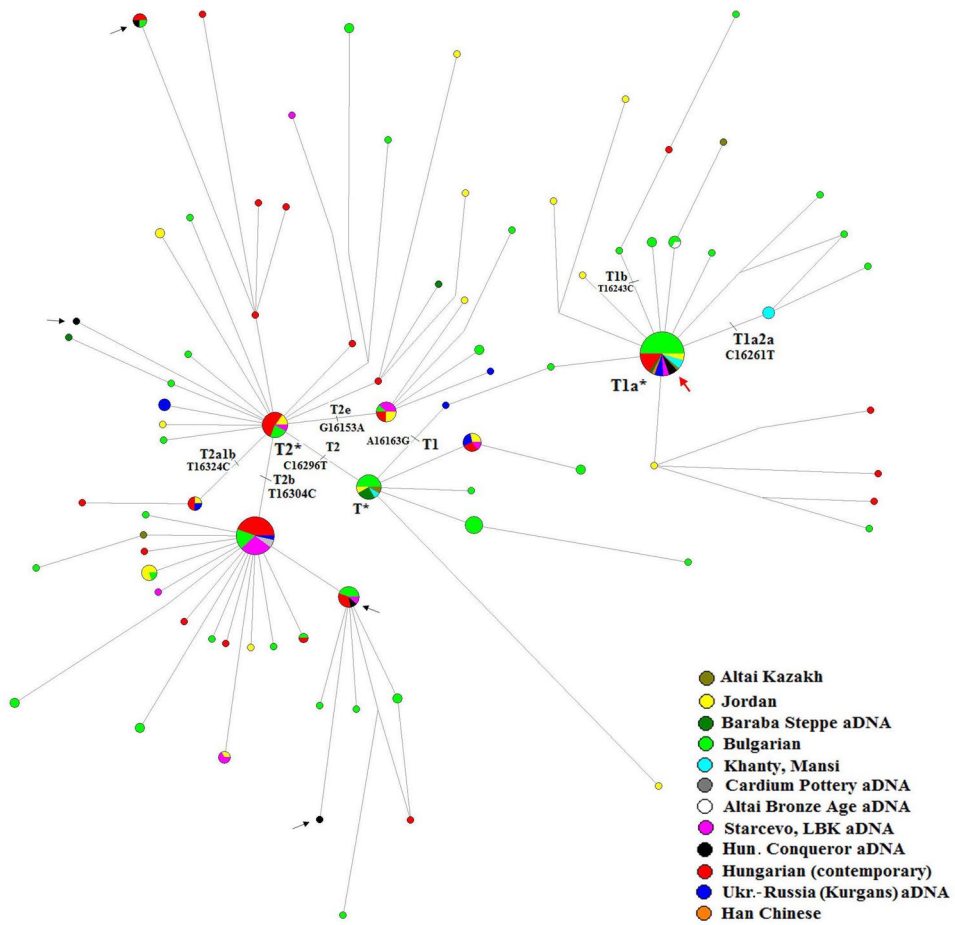

Fig. 6 Median-Joining network for Haplogroup T. The *circle sizes* are proportional to the haplotype frequencies; the smallest area is equivalent to one individual. *Colours* indicate populations, as listed at the *bottom* of the figure. *Red arrow* shows KAR individuals, and *black arrows* show AH2 individuals. The published haplotypes from different populations used for the network are presented in Online Resource 4 (ESM_4) (colour figure online)

belonged to *T2b* (not nodal) and *T2*(T2b,T2e,T2a1b)* haplotypes (black arrows in Fig. 6). As *T1a* has significant South Siberian presence (both contemporary and Bronze Age), while *T2b* is absent from Asia but shared by both Danubian and Mediterranean Neolithic aDNA samples, we reason that the Karos conqueror *T1a* samples rather have South Siberian maternal roots, while the previously published *T2* samples from the tenth century Carpathian Basin more likely represent haplotypes of Neolithic European origin.

Median-Joining network for Haplogroup *B* is shown in Fig. 7. The founder *B4* haplotype was shared by 4 samples: 2 Karos conquerors (red arrow in Fig. 7), 1 Han Chinese from Liaoning, and 1 Uzbek from Tashkent. The two Karos *B4′5* samples (blue arrow in Fig. 7) were at two mutation distances from the founder haplotype. An earlier published Hungarian conqueror *B* haplotype (black arrow on Fig. 6) matched one Altai Kazakh from the *B4c1b* subgroup. One living Hungarian sample was at one mutation distance from this conqueror. No living Hungarians were close to the Karos conqueror *B* samples.

As Liaoning was historically inhabited by Tungus-speaking Manchu, thus considering the Uzbek and Altai Kazakh conqueror match, we assume that mtDNA Hg *B* among conquerors is most likely an Altaic (Turkic) ethnic signal.

Y-chromosome haplogroup data

We could amplify Y-chr SNPs (ESM_2) and STRs (ESM_3) with a limited success, due to the low copy number of nuclear loci. In spite of our optimization efforts, some of the multiplex primers worked poorly; namely, *M17* was absent from all samples, and *M35* gave weak signal even from modern DNA. Moreover, *M172* gave strong background signal in each sample, indicating background contamination, however, as its major haplogroup *M304* showed the ancestral allele in all cases, the *M172* peak rather arose from non-human background, and probably for similar reasons *M216* also gave occasional false signal. We were able to supplement some of the missing or ambiguous signals via optimized single-plex PCR experiments (yellow boxes in ESM_2), which confirmed and refined the GenoY25 multiplex results. Finally, eliminating the *M172* background and interpreting all signals from each extract on the phylogenetic tree, we could decipher the Hg of four

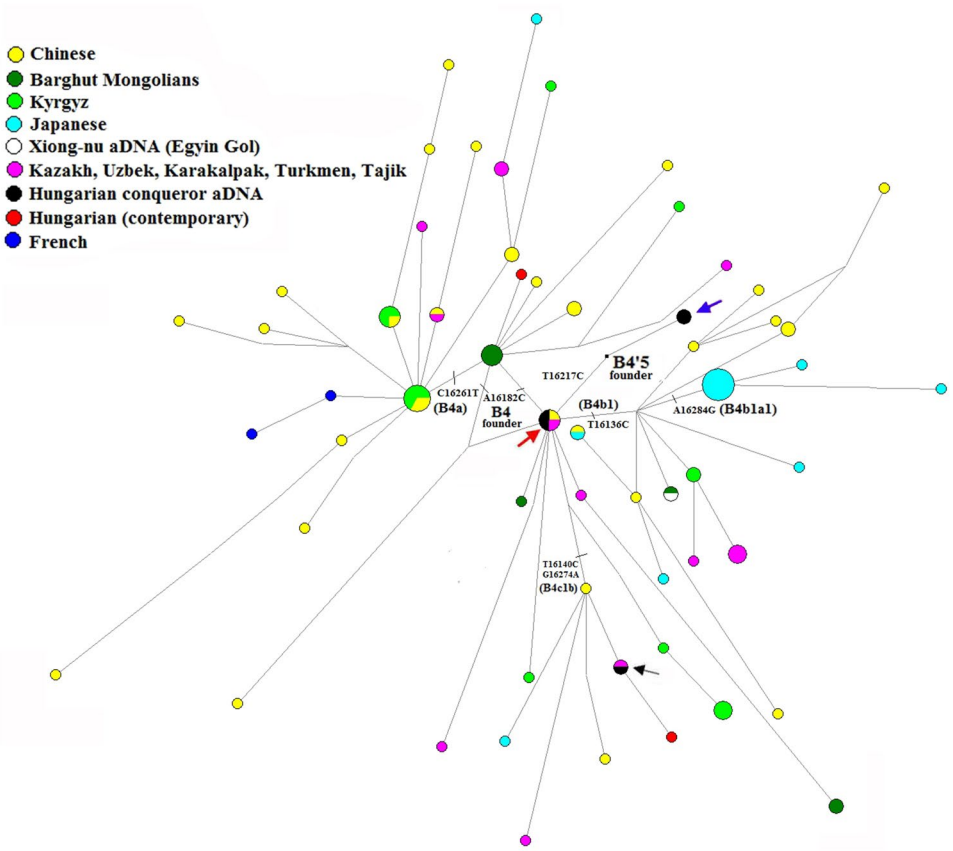

Fig. 7 Median-Joining network for Haplogroup B. The *circle sizes* are proportional to the haplotype frequencies; the smallest area is equivalent to one individual. *Colours* indicate populations, as listed at the top of the figure. *Red* and *blue arrows* show KAR individuals, and *black arrow* shows AH2 individual (colour figure online)

samples. Samples 1 and 3 belong to Hg *R1b1b1a*, while samples 12 and 17 to Hg *I*. It is notable that Hg-s of samples 12 and 17 were also confirmed by independent Y-chromosomal STR data (ESM_3), which further refined their classification into sub-Hg *I2a,* although haplogroup prediction using Y-STR profiles may be error-prone (Wang et al. 2015). As typical for aDNA, our STR amplification success decreased with product length (see Online Resource 3 (ESM_3) and we could obtain only fragmentary STR profiles with loci or allele dropouts. We need to add that the STR method requires much DNA, which used up our best extracts (typically from teeth), so we could not repeat these experiments.

It is interesting that paternal Hg-s of nearly one-third of the males were deciphered, but merely, typical European Y Hg-s (*I2a* and *R1b*) were detected. Hg *I-M170* is predominantly European and it is considered as the only native European Y Hg (Rootsi et al. 2004), which is present in the Carpathian Basin since the Neolithic (Gamba et al. 2014), and missing from Asia. *I2a* has a high, 16.74% frequency among modern Hungarians. *R1b-M269* is the most frequent Hg in Western Europe, and despite of its Asian origin, today occurs there only sporadically (Myres et al. 2011). *R1b* is found in 18.1% of modern Hungarian males. Its subgroup *R1b-S21 (U106)* is typical among German-speakers, and is absent from Asia. Our data suggest that a considerable proportion of the paternal lineages of the Karos population were derived from Europe.

Autosomal STR data

We also tried to decipher kinship relations in the Karos III. graveyard by autosomal STR analyses. Identical mtDNA haplotypes of samples 1 and 5 (Hg *B4*), samples 6 and 13 (Hg *B4′5*), samples 3 and 17 (Hg *H6*), and samples 18 and 19 (Hg *T1a10*) suggested that these individuals could have been directly related maternally. Of these possibilities, our fragmentary STR data (ESM_3) could exclude the direct relation of samples 3 and 17, since three of their examined loci (*D21S11, D2S1338,* and *VWA*) showed different heterozygote genotypes which exclude allele dropouts and confirm exclusion of close relatives. In contrast, we could confirm the possible relation of samples 18 and 19, since five of the loci (*D3S1358, D10S1248, D16S539, D2S441,* and *D12S391*) amplified from both samples shared the same alleles. The other two relations remained obscure due to poor DNA quality.

Possible paternal relations of samples 12 and 17 with identical haplogroups (*I2a*) could also be excluded by both autosomal (*D8S1179* and *D12S391*) and Y-chr (*DYS390*)

STR data. It is notable, however, that STR data of samples 1 and 3 show a strong indication of direct paternal relations with ten shared loci (*D3S1358, TH01, D21S11, D10S1248, D1S1656, D2S1338, D22S1045, VWA, D2S441,* and *D12S391*). As according to the GenoY25 assay sample 1 and 3 falls in the same haplogroup *R1b1b1a* (ESM_2) and both died at an old age (Fig. 1), these individuals were likely brothers.

Discussion

In the last 10 years, ancient DNA research has significantly transformed our knowledge about the early peopling of Europe (reviewed in Haber et al. 2016; Brandt et al. 2015; Novembre 2015), but data following the Bronze Age are sporadic, though historically documented migrations further transformed the continent in a great extent. Our results provide new data about one of these later migrations, which can facilitate to piece together the genetic changes having shaped the continent after the Bronze Age. We have shown that the Hungarians redelivered haplogroups already present on the continent, but also refreshed the European gene pool with new Asian genes.

The maternal genetic composition of the Karos conqueror population indicates a mixed origin from multiple sources. About 30–40% of the population is obviously originated from East-Middle Asia, at least on a wider time scale, and our Hg *B* network data indicate that this component might have derived from a common gene pool with the today Altaic (Turkic) speaking ethnic groups. In spite of the considerable proportion of Asian Hg-s, present in the conquerors, their anthropological characters were predominantly Europid, with few Europo-Mongoloid individuals (Éry 1994), and the same is true for the Karos III. cemetery (Kustár 1996). This suggests that they acquired the Asian admixture generations before the conquest on the East-European Pontic steppes. In a search for ancient populations on this area with similar Hg distributions to KAR, our IRC and SOC analyses indicate that the Sintashta–Andronovo (SIA) and Late Baraba (BB3) are the youngest ancient cultures closest to KAR (Fig. 4 MDS plot), falling into the intersection of Western and Siberian gene flow. The geographical location of SIA and BB3 populations fits into historical expectations, and they already contained a balanced composition of Western and Siberian CHgC-s (Molodin et al. 2012). This suggests that the admixture founding a basic component of the conqueror gene pool may have happened in Central Asia-South Siberia more than 2000 years before the conquest, around the SIA and BB3 horizon. In our SOC analysis (ESM_5) Western and Siberian components are most balanced in clusters 2 and 3, and their Western components are also more similar to each other than to cluster 1, which is indicative of genetic relations between members of these clusters. As we have shown with SOC analyses, the closest contacts of KAR are found among populations of clusters 2 and 3 (Fig. 5b).

It is notable that cluster 2 contains all Hungarian ancient and recent populations: AH1 (Ancient Hungarian), AH2 (Hungarians 900 AD), SEK (Sekler), and CSG (Csángó), except HUB (Hungarian Budapest). This is reasonable, since haplogroups of Asian origin (*A, B, C,* and *D*) occur below 1% frequency in the current Hungarian, and above 1% frequency in the Hungarian speaking Sekler and Csángó populations (latter two live today in Romania), and latter groups also contain relatively high, 5–7.4% Central and Inner Asian paternal haplogroups (Bíró et al. 2015), distinguishing them from Western European populations. In agreement with SOC analyses, our Hg T network data also indicate, that in contrast to the conclusions of Tömöry et al. 2007, genetic continuity can be detected between ancient and modern Hungarians. It is also remarkable that 14 of the 35 ancient populations are clustered in cluster 2 (ESM_5), and this number increases to 24 in the unified population set of clusters 2 and 3 containing quite all ancient population found in the Carpathian Basin LBT, STR, SZA, LOM, SLV, AH1, AH2, and KAR. A possible explanation of this unexpected result can hide in the Middle East Neolithic (MEN) population, which is not far from KAR on the MDS map in Fig. 5b, and may indicate that all these populations contain considerable genetic input from the Early Neolithic Fertile Crescent.

From a linguistic point of view, we should state that very few recent Indo-European speakers (SCI = Scottish and KUR = Kurdish) are found among the recent populations in the unified set of clusters 2 and 3, whereas the majority belong to Turkic, Mongolian, Uralic, and Kartvelian language groups (see ESM_4).

We need to stress that a significant portion of the Karos population cannot have derived from Siberia, since the *X2F* and *H5* maternal, and the *I2a* and *R1b1b* paternal lineages are rare or absent there. The presence of Hg-s *X2f* and *H5*, the relatively high frequencies of *H6* and *U2* may indicate a genetic connection of the KAR population with the Caucasus region, where the *T1a* and *J* Hg-s are also known to be present in above average frequencies. The above-mentioned archaeological parallels with the Martan-Ču findings also support the Caucasus connection. Unfortunately, ancient DNA data are sporadic from this region, and in our modern data set, the Caucasus is also underrepresented. These data indicate that the Hungarians assimilated various further populations along their way to Europe, in agreement with our historical knowledge about tribal organizations. This is also supported by the fact that although all conqueror populations were clustered together, the KAR population shows a significant difference from the AH1 and AH2 conqueror

populations, which were unearthed from different cemeteries, and have higher European components. It is possible that genetic differences mirror different tribal origin.

We could also reveal that the Karos III. graveyard contains family members, and despite the small number of individuals, at least two pairs of direct relatives could be identified. Of these, the no. 18 young woman was probably the mother of the no. 19 boy, as they were buried next to each other. The no. 1 and no. 3 old men were probably brothers, contradicting their archaeological classification, according to which no. 3 was a warrior, while no. 1 could have been servant.

To further clarify the relation of the ancient Hungarians to previous and later inhabitants of the Carpathian Basin as well as to contemporary populations on the Pontic steppes, further aDNA data are needed from these regions and periods. Hitherto most aDNA research was focused on the peopling of early Eurasia from the Mesolithic to the late Bronze Age, and data from later periods are underrepresented. We intend to re-examine all three Karos graveyards and other conqueror burials by determining the whole mtDNA genomes as well as Y-chr haplogroups with High-Throughput Sequencing. This would also enable us to ascertain the exact kinship relations within and between cemeteries, as well as to illuminate the genetic structure of their tribal organization.

Acknowledgement The generous support of Avicenna Foundation Grant No. GF/JSZF/814/9/2015 to I.R. and encouragement of Professor Miklos Maroth is highly appreciated. This research was also supported in part by OTKANN 78696, OTKA K78555, Hungary-Romania Cross-Border Co-operation Programme No. HURO/1101/173/2.2.1, and Campus Hungary grant B2/1SZ/18826. Niall O´Sullivan is highly acknowledged for his technical support in the ancient DNA laboratory.

Compliance with ethical standards

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

Bartram AK, Poon C, Neufeld JD (2009) Nucleic acid contamination of glycogen used in nucleic acid precipitation and assessment of linear polyacryl amide as an alternative co-precipitant. Biotechniques 47:1019–1022

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Benoit J-N, Quatrehomme GF, Carle G, Pognonec P (2013) An alternative procedure for extraction of DNA from ancient and weathered bone fragments. Med Sci Law 53:100–106

Bíró A, Fehér T, Bárány G, Pamjav H (2015) Testing Central and Inner Asian admixture among contemporary Hungarians. Forensic Sci Int Genet 15:121–126

Bogácsi-Szabó E, Csányi B, Tömöry Gy, Blazsó P, Csősz A, Kiss D, Langó P, Köhler K, Raskó I (2008) Archeogenetikai vizsgálatok a Kárpát-medence 10 századi népességén Magy Tudomány 10:1204–1217

Brandstatter A, Egyed B, Zimmermann B, Duftner N, Padar Z, Parson W (2007) Migration rates and genetic structure of two Hungarian ethnic groups in Transylvania Romania. Ann Hum Genet 71:791–803

Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, Haak W (2015) Human paleogenetics of Europe : the known knowns and the known unknowns. J Hum Evol 79:73–92

Champlot S, Berthelot C, Pruvost M, Andrew Bennett E, Grange T, Geigl EM (2010) An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PLoS One. doi:10.1371/journalpone0013042

Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME (2015) Improving access to endogenous DNA in ancient bones and teeth. Sci Rep 5:11184. doi:10.1038/srep11184

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kővári I, Pap I, Anders A, Whittle A, Dani J, Raczky P, Higham TFG, Hofreiter M, Bradley DG, Pinhasi R (2014) Genome flux and stasis in a five millennium transect of European prehistory Nat Commun 5:5257. doi:10.1038/ncomms6257

Haak W, Brandt G, Jong HND, Meyer C, Ganslmeier R, Heyd V, Hawkesworth C, Pike AWG, Meller H, Alt KW (2008) Ancient DNA Strontium isotopes and osteological analyses shed light on social and kinship organization of the Later Stone Age. PNAS 105:18226–18231

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Villems R, Meller H, Alt KW, Cooper A (2010) Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. PLoS Biol. doi:10.1371/journalpbio1000536

Haber M, Mezzavilla M, Xue Y, Tyler-Smith C (2016) Ancient DNA and the rewriting of human history: be sparing with Occam's razor. Genome Biol 17:1. doi:10.1186/s13059-015-0866-z

Juhász Z, Fehér T, Bárány G, Zalán A, Németh E, Pádár Z, Pamjav H (2015) New clustering methods for population comparison on paternal lineages. Mol Genet Genomics 290:767–784

Juhász Z, Fehér T, Németh E, Pamjav H (2016) mtDNA analysis of 174 Eurasian populations using a new iterative rank correlation method. Mol Genet Genomics 291:493–509

Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA (2012) Setting the stage: Building and working in an ancient DNA laboratory. Ann Anat 194:3–6

Kustár Á (1996) A karosi honfoglaláskori temetők in: Kovács L Révész L (eds) Herman Ottó Múzeum és Magyar Nemzeti Múzeum Miskolc pp 395–402

Lahermo P, Laitinen V, Sistonen P, Béres J, Karcagi V, Savontaus ML (2000) mtDNA polymorphism in the Hungarians: comparison to three other Finno-Ugric-speaking populations. Hereditas 132:35–42

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Sudmant PH, Schraiber JG, Castellano S, Kirsanow K, Economou C, Bollongino R, Fu Q, Bos K, Nordenfelt S, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Ayodo G, Babiker H, Balanovska E, Balanovsky O, Ben-Ami H, Bene J, Berrada F, Brisighelli F, Busby GBJ, Cali F, Churnosov M, Cole DEC, Damba L, Delsate D, van Driem G, Dryomov S, Fedorova S, Francken M, Gallego Romero I, Gubina M, Guinet J-M, Hammer M, Henn B, Helvig T, Hodoglugil U, Jha AR, Kittles R, Khusnutdinova E, Kivisild T, Kučinskas V, Khusainova R, Kushniarevich A, Laredj L, Litvinov S, Mahley RW, Melegh B, Metspalu E, Mountain J, Nyambo T, Osipova L, Parik J, Platonov F, Posukh OL, Romano V, Rudan I, Ruizbakiev R,

Sahakyan H, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Voevoda M, Wahl J, Zalloua P, Yepiskoposyan L, Zemunik T, Cooper A, Capelli C, Thomas MG, Tishkoff S, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413

Molodin VI, Pilipenko AS, Romaschenko AG, Zhuravlev AA, Trapezov RO, Chikisheva TA (2012) Human migrations in the southern region of the West Siberian Plain during the Bronze Age: Archaeological palaeogenetic and anthropological data in: Population Dynamics in Prehistory and Early History: New Approaches Using Stable Isotopes and Genetics pp 93–112 (Eds: Kaiser E, Burger J and Schier W.) doi:10.1515/978311026630693

Myres NM, Rootsi S, Lin A, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, Balanovsky O, Balanovska E, Rudan P, Baldovic M, Herrera RJ, Chiaroni J, Di Cristofaro J, Villems R, Kivisild T, Underhill P (2011) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet 19:95–101

Novembre J (2015) Ancient DNA steps into the language debate. Nature 522:164–165

Pamjav H, Juhász Z, Zalán A, Németh E, Damdin B (2012) A comparative phylogenetic study of genetics and folk music. Mol Genet Genomics 287:337–349

Pamjav H, Juhász Z, Fehér T, Bárány G, Németh E (2013) Classification of the Y-haplogroup distributions of Western Eurasian populations using a self-learning algorithm Forensic Sci Int Genet Suppl Ser 4 doi:10.1016/jfsigss201310101

Pletneva SA (1981) Saltovo-maiatskaia kultúra in: Pletneva SA (Ed) Stepi Evrazii v Epohu Srednevekov'ia Moskva pp 62–75

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H, Parik J, Loogväli E-L, Derenko M, Malyarchuk B (2003) Origin and Diffusion of mtDNA Haplogroup X. Am J Hum Genet 73:1178–1190

Révész L (1996) A karosi honfoglaláskori temetők. Herman Ottó Múzeum és Magyar Nemzeti Múzeum Miskolc

Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. Nat Protoc 2:1756–1762

Róna-Tas A (1999) Hungarians and Europe in the early middle ages: an introduction to early hungarian history. CEU Press

Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R (2007) Origin and expansion of haplogroup H the dominant human mitochondrial DNA lineage in west Eurasia: the Near Eastern and Caucasian perspective. Mol Biol Evol 24:436–448

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barać L, Pericić M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioğlu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan J-P, Chaventre A, Ferak V, Füredi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzić R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am J Hum Genet 75:128–137

Szécsényi-Nagy A, Brandt G, Haak W, Keerl V, Möller-Rieker S, Köhler K, Mende BG, Oross K, Marton T, Osztás A, Kiss V, Fecher M, Pálfi G, Molnár E, Sebők K, Czene A, Paluch T, Šlaus M, Novak M, Pećina-Šlaus N, Ősz B, Voicsek V, Somogyi K, Tóth G, Kromer B, Bánffy E, Alt KW (2015) Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. Philos Trans R Soc B Biol Sci 282:20150339. doi:10.1098/rspb20150339

Tömöry G, Csányi B, Bogácsi-Szabó E, Kalmár T, Czibula Á, Csősz A, Priskin K, Mende B, Langó P, Downes CS, Raskó I (2007) Comparison of maternal lineage and biogeographic analyses of ancient and modern Hungarian populations. Am J Phys Anthropol 134:354–368

Vinogradov VB (1983) Altungarische Parallelen zu einigen Gräbern des alanischen Gräberfeldes bei Martan-Ču. Acta Archaeol Hungarica 35:211–220

Wang CC, Wang LX, Shrestha R, Wen S, Zhang M, Tong X, Jin L, Li H (2015) Convergence of Y Chromosome STR haplotypes from different SNP haplogroups compromises accuracy of haplogroup prediction. J Genet Genomics 47:403–407