

A study of the Bodrogek population in north-eastern Hungary by Y chromosomal haplotypes and haplogroups

Horolma Pamjav¹ · Á. Fóthi^{2,3} · T. Fehér⁴ · Erzsébet Fóthi⁵

Received: 23 January 2017 / Accepted: 9 April 2017 / Published online: 13 April 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract We have determined the distribution of Y chromosomal haplotypes and haplogroups in population samples from one of the most important areas in north-eastern Hungary from many villages in the Bodrogek. The Bodrogek region was chosen due to its isolated nature, because this area was a moorland encircled by the Tisza, Bodrog, and Latorca Rivers and inhabitants of this part of Hungary escaped from both Tatar and Ottoman invasions, which decimated the post-Hungarian Conquest populations in many parts of the country. Furthermore, in the first half of the tenth century, this region served as the Palatial Centre and burial grounds of the Hungarian tribes. It has thus been assumed that the present population in this area is likely to be more similar to the population that lived in the Conquest

period. We analysed male-specific markers, 23 Y-STRs and more than 30 Y-SNPs, that reflect the past and recent genetic history. We found that the general haplogroup distribution of the samples showed high genetic similarity to non-Bodrogek Hungarians and neighbouring populations, despite its sheltered location and historical record. We were able to classify the Y-chromosomal haplogroups into four large groups based on STR mutation events: pre-Roman/Roman ancient lineage, Finno-Ugric speakers arriving into the Carpathian Basin, Migration period admixture, and post-Hungarian Conquest admixture. It is clear that a significantly larger database with deep haplogroup resolution, including ancient DNA data, is required to strengthen this research.

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-017-1319-z) contains supplementary material, which is available to authorized users.

✉ Horolma Pamjav
phorolma@hotmail.com

✉ Erzsébet Fóthi
fothi.erzsebet@nhmus.hu

¹ National Centre of Forensic Experts and Research, Budapest, Hungary

² Department of Genetics, Faculty of Sciences, Eötvös Loránd University, Budapest, Hungary

³ Research Centre for Natural Sciences, Institute of Enzymology, Budapest, Hungary

⁴ The Hungarian Magyar Family Tree DNA Project, Budapest, Hungary

⁵ Department of Anthropology, Hungarian Natural History Museum, Budapest, Hungary

Keywords Y haplogroup and Y haplotype analysis · Phylogenetic study · Recent Bodrogek Hungarian population · Hungarian ancient history

Introduction

Y-STRs and Y-SNPs on the non-recombining region of the human Y chromosome have been widely studied to understand the population history, origin, and migration of human populations (de Knijff 2000; Jobling and Tyler-Smith 2003). Such studies have greatly benefited from analyses of the non-recombining region of the Y chromosome, as this region is not altered by the conventional recombinational processes and is transmitted intact from one generation to the next, revealing the deep paternal lineages of a population. Populations share Y-chromosome lineages because of common origins or gene flow from the populations in neighbouring areas.

Hungarian Y-chromosome studies have been analysed (Völgyi et al. 2009; Bíró et al. 2015; Csányi et al. 2008), but only for 12 STRs and without a regional focus (except for Székely and Csángó populations from Romania). We have, therefore, analysed a specific Hungarian sub-population for 23 Y-STRs. The Bodroghköz region was chosen due to its isolated nature. Until the end of the 19th century, this area was a moorland encircled by the Tisza, Bodrog, and Latorca Rivers. Living in *refugium*, inhabitants of this part of Hungary escaped from both Tatar and Ottoman invasions, which decimated the post-Hungarian Conquest populations in many parts of the country. Furthermore, in the first half of the tenth century, this region served as the Palatial Centre and burial grounds of the Hungarian tribes (Révész 1996). It has thus been assumed that the present population in this area is likely to be more similar to the population that lived in the Hungarian Conquest period than to less isolated and protected populations in other areas of Hungary (Csányi et al. 2008).

Materials and methods

Materials

We collected 147 unrelated male samples from 24 small villages (Fig. 1) in Bodroghköz in north-eastern Hungary. Informed consent was obtained from all individual participants included in the study. The samples published herein were sent to the YHRD and the accession number is YA004249.

Methods

Testing of Y-STR and Y-SNP markers

Genomic DNA was extracted from buccal swab samples using the DNAIQ system (Promega, Madison, WI, USA), as described in the manufacturer's instructions. The samples were quantified using the ABI 7500 Real-time PCR System (Life Technologies, Foster City, CA, USA).

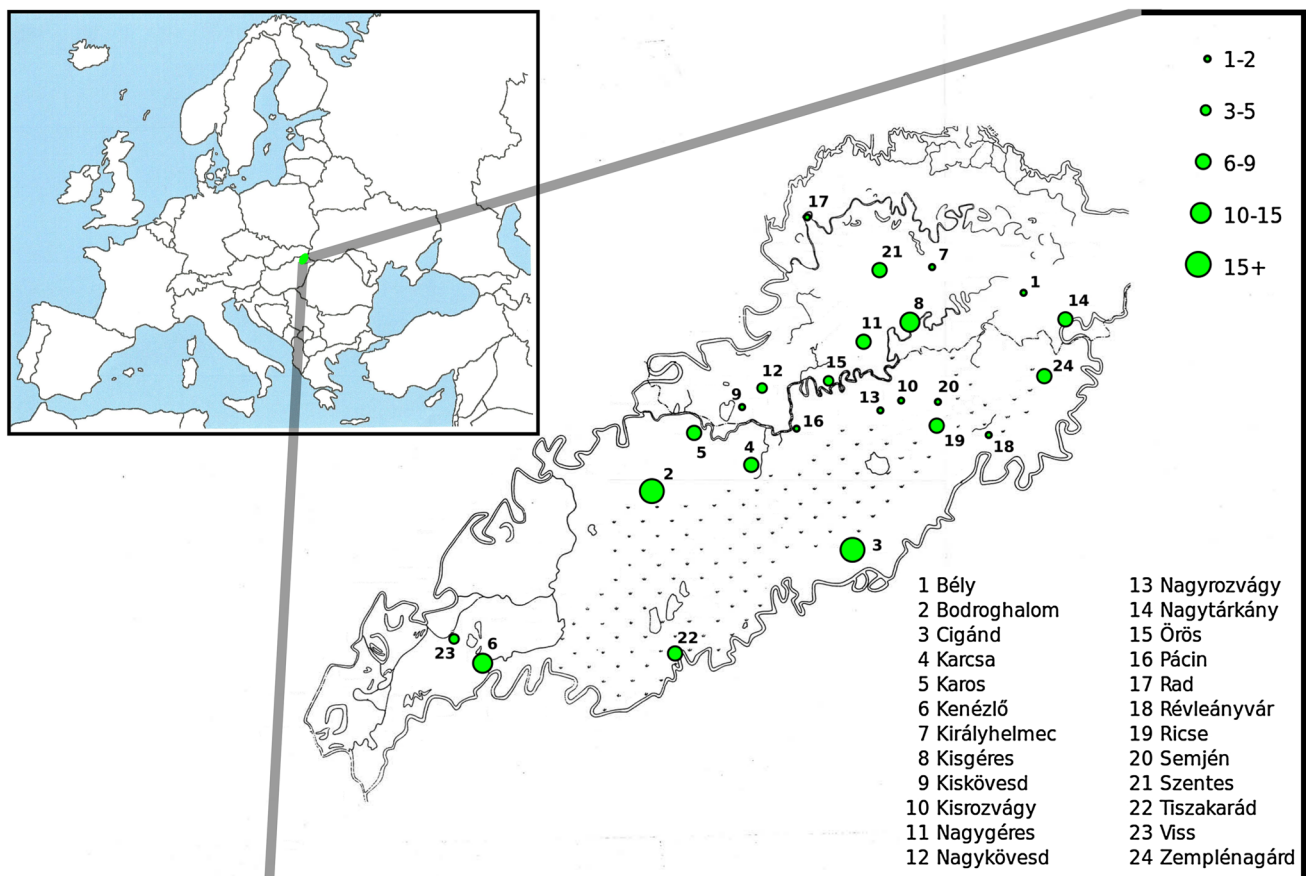


Fig. 1 147 non-related Hungarian male individuals representing 24 different geographical regions in Bodroghköz of North-Eastern Hungary. The circle sizes are proportional to the sample sizes

DNA was amplified using the Promega PowerPlex Y23 amplification kit, including 23 Y-STR loci, according to the manufacturer's instructions. Fragment sizes and allele designations were determined using a Genetic Analyzer ABI3130 (Applied Biosystems, Foster City, CA, USA) using the GeneMapper ID-X v.1.2 software. When testing Y-SNP markers, amplifications of 3–5 ng genomic DNA were performed in an ABI 7500 and in GeneAmp 9700 thermal cyclers, with TaqMan probes using the programs designed by ABI. The relative fluorescence of the PCR products was analysed on an ABI 7500 Real-time PCR System using its SDS software, as described in the ABI manufacturer's manual. The SNP markers tested were E-M35, E-M78, E-M81, E-M123, G-M201, G2-L156, G1-M285, H1-M52, H1a-M82, I-M170, I1-M253, I1-L22, I2a1-M223, I2a2-P37, J-M304, J1-M267, J2-M172, J2-M67, J2b-M12, R1a-M198, R1a-M458, R1a-Z93, R1a-Z280, R-M207, R1-M173, R2-M124, R1b-M343, R1b-M269, R1b-U106, R1b-P312, R1b-M412, N-M231, N1c-L1034, N1c-VL29, N1c-Z1936, and Q-M242.

A list of primers and TaqMan assays for binary markers is used in the laboratory, as shown in Online Resource 1 (ESM_1).

Phylogenetic study

To examine the STR variation within the sub-haplogroups, relational networks were constructed using the Network 5.0.0.0 program (Bandelt et al. 1999). Repeats of the DYS389I locus were subtracted from the DYS389II locus. To put the results into a more extensive geographical context, we included haplotypes of 17 overlapping loci from other populations. Within the network programme, the rho statistic was used to estimate the time to the most recent common ancestor (TMRCA) of haplotypes within the compared haplogroups (Bandelt et al. 1999). We used the pedigree rate 2.5×10^{-3} , as described by Goedbloed et al. (2009), for 17 loci included in the YFilter kit (Applied Biosystem, Foster City, CA, USA). Due to the expansion of Full Genomic testing, an SNP count-based TMRCA estimation method was also used and proved to be effective on closely related, as well as genetically distant individuals (Hallast et al. 2015; Underhill et al. 2015). In these recent studies, the pedigree rate based on large number STR loci (Goedbloed et al. 2009) tended to be significantly closer to the SNP-based time estimates.

Genetic structure

Pairwise F_{st} genetic distances were calculated based on haplogroup frequencies using the Arlequin 3.5 software (Excoffier and Lischer 2010). Y-STR-based pairwise R_{st} genetic distances were computed using YHRD.ORG's online AMOVA program (release 52) (Willuweit and Roewer 2015).

Kruskal's non-metric multidimensional scaling (MDS) plots were constructed with the isoMDS() function from the R package 'MASS' (ver. 7.3-45) (Venables and Ripley 2002).

Results

The frequencies of the haplogroups are presented in Table 1. The 23 Y-STR haplotype and haplogroup results of the 147 males are shown in Online Resource 2 (ESM_2). The largest haplogroups were R1a-M458 (20.4%), I2a1-P37 (19%), R1a-Z280 (14.3%), and E1b-M78 (10.2%). Various R1b-M343 subgroups accounted for 15% of the Bodrogköz population. Haplogroup N1c-Tat covered 6.2% of the lineages, but most of it belonged to the N1c-VL29 subgroup, which is more frequent among Balto-Slavic speaking than Finno-Ugric speaking peoples. Other haplogroups had frequencies of less than 5%.

Phylogenetic study

Based on Y-STR loci (haplotypes), networks were constructed within each of the haplogroups. To put the results into a wider geographical context, we included

Table 1 Haplogroup distribution of the Bodrogköz population group

Haplogroup	<i>N</i>	%
E-M78	15	10.2
E-M81	1	0.7
G2a-L156	3	2.0
H1a-M82	2	1.4
I1-M253	7	4.8
I2a1-P37	28	19.0
I2a2-M223	1	0.7
J1-M267	1	0.7
J2b-M12	5	3.4
N1c-M178*	1	0.7
N1c-VL29	6	4.1
N1c-Z1936*	1	0.7
N1c-L1034	1	0.7
Q-M242	1	0.7
R1a-M458	30	20.4
R1a-Z280	21	14.3
R1b-M343 (xM412)	6	4.1
R1b-M412	1	0.7
R1b-P312	11	7.5
R1b-U106	4	2.7
R2-M124	1	0.7
Total	147	100.0

haplotypes of 17 overlapping loci from other populations. We included three networks (N1c-VL29, J2b-M12, and R1a-M458) in the main text of the study that could be potentially useful for unravelling the genetic history of Hungarians. Other networks belonging to non-frequent subgroups are not included in this study.

A median-joining network of 151 N1c-VL29 haplotypes was generated using Bodrogeköz Hungarians ($n = 5$), Ukrainians ($n = 33$), Belarusians ($n = 22$), Central Russians ($n = 14$), Northern Russians ($n = 6$), Slovaks ($n = 4$), Lithuanians ($n = 17$), Estonians ($n = 26$), Finns ($n = 5$), Saami ($n = 5$), Karelians ($n = 5$), Vepsa ($n = 1$), Mari ($n = 5$), Komi ($n = 1$), and Chuvash ($n = 3$). DYS635 and DYS385a-b (3 loci) data were not available in the N1c-Tat network published by Ilumae et al. (2016). The remaining 14 loci had an estimated 657–788 year/1 molecular step (25 vs. 30 year/generation), based on pedigree data. We compared the data with SNP-based estimates available at YFull (<https://www.yfull.com/tree/>).

TMRCA estimates (with 30 years/generation):

All VL29 ($n = 151$): 2541 ± 473 years (YFull by SNPs: 3500 ± 500 years).

M2783 (“Balto-Slavic” branch): 2373 ± 561 years (YFull: 2800 ± 400 years).

The founder N1c-VL29 haplotype was shared by 11 samples (4 Estonians, 3 Karelians, 3 Northern Russians, and 1 Komi), which is in agreement with the postulated Finno-Ugric origin of the clade (Fig. 2, see cluster 1). We did not test our samples deeper than N1c-VL29; however, based on deep SNP data from public databases, we have further divided into subgroups. There is an N1c-M2783 (whose new name is N3a3-M2783) subgroup defined by DYS19 = 15 (instead of DYS19 = 14), which is the predominant N1c-VL29 subgroup among Balto-Slavic-speaking populations (Fig. 2, clusters 2, 3, 4, 6). The founder haplotype was a Lithuanian male (Fig. 2, haplotype 2). N1c-M2783 (xZ16981) modal (Fig. 2, cluster 6) was the second most populous haplotype, with 9 haplotypes (5 Ukrainians, 2 Lithuanians, 1 Belarusian, and 1 Hungarian sample from Bodrogeköz). At one molecular step, the L551 (Fig. 2, cluster 4) founder haplotype (notably the clade of Gedeminids, i.e., Lithuanian Gedeminid and Polish Jagiellonian kings) was shared by two Lithuanians, two Ukrainians, one Belarusian, and one Central Russian. The Z16981 (Fig. 2, cluster 3) founder haplotype contained 1–1 Estonian, Slovak, Belarusian and Central Russian. L1022 (Fig. 2, cluster 5) consisted of predominantly Baltic-Finnic haplotypes, but no Hungarians. It should also be noted that no Hungarian belonged to the L551 clade, which means that we could not anchor the genetic descendants of Jagiellonians in Bodrogeköz. To summarize, out of five Bodrogeköz Hungarian N1c-VL29 samples, four may be of post-Hungarian Conquest

Lithuanian or Slavic admixture and one of Finno-Ugric origin, with no recent matches.

A median-joining network of 175 R1a-M458 samples was generated with Bodrogeköz Hungarians ($n = 30$), Slovaks ($n = 57$), Slovak Roma ($n = 7$), Poles ($n = 23$), Czechs ($n = 6$), Germans ($n = 13$), Dutch ($n = 2$), Danish ($n = 1$), Austrians ($n = 1$), Croatians ($n = 14$), Bosnians ($n = 3$), Slovenians ($n = 1$), Macedonians ($n = 1$), Ukrainians ($n = 7$), Russians ($n = 5$), Konda Mansi ($n = 2$), and Estonians ($n = 2$), taken from data published by Underhill et al. (2015). This network had two large clusters. We have not tested SNPs below M458; however, based on deep SNP data from public databases, they can be defined by SNPs L260 and L1029. The L1029 founder haplotype was shared by two Hungarians and three Slovaks (Fig. 3, see cluster 1). In the L260 area (Fig. 3, see cluster 2), two Hungarian and one Estonian formed the founder haplotype. Based on a pairwise comparison of all haplotypes, eight samples are ancient (Migration period) Slavic admixture, two are likely Finno-Ugric, six Bodrogeköz samples may be a post-Hungarian Conquest Slovak/Polish admixture, two are likely Finno-Ugric, and eight are a more ancient Slavic admixture (from the Migration period), while 14 R1a-M458 samples are unlikely to be of recent Slavic admixture and are of unspecified origin. It is important to note that Mansi R1a-M458 samples do not have close Russian or other Slavic matches, but they may be of recent Slavic admixture. TMRCA calculations based on 17 STR loci (556–668 years/mutation) were very close to SNP-based estimates of YFull.

TMRCA estimates (30 years/generation):

L1029 area: 2063 ± 327 YBP (YFull: 2100 ± 200 YBP).

L260 area: 2442 ± 368 YBP (YFull: 2600 ± 500 YBP).

A median-joining network of 65 J2b-M12 haplotypes was generated with Bodrogeköz Hungarians ($n = 5$), Serbians ($n = 3$), Bulgarians ($n = 12$), Ukrainians ($n = 2$), Austrians from East Tyrol ($n = 4$), Northern Albanians/Gheg ($n = 30$), and Central Asians ($n = 9$) (Sarno et al. 2016; Mirabal et al. 2009; Karachanak et al. 2013; Mielnik-Sikorska et al. 2013; Di Cristofaro et al. 2013; Niederstätter et al. 2012). This was probably the most surprising of all the networks. While most Albanians and Bulgarians clustered together, Serbian, Hungarian, and Austrian J2b-M12 haplotypes were very distant from the South Balkan cluster and in some cases were closer to Asian haplotypes. Four Hungarian samples split from a common ancestor much earlier (as illustrated by the diversity and the molecular mutational steps shown in Fig. 4), and only one belonged to the Albanian–Bulgarian cluster.

TMRCA estimates (30 years/generation):

Albanian–Bulgarian cluster: 1781 ± 381 YBP.

Serbian–Hungarian–MidEast branch: 5344 ± 1215 YBP.

YFull (SNP-based) all M12: $15,700 \pm 1700$ YBP.

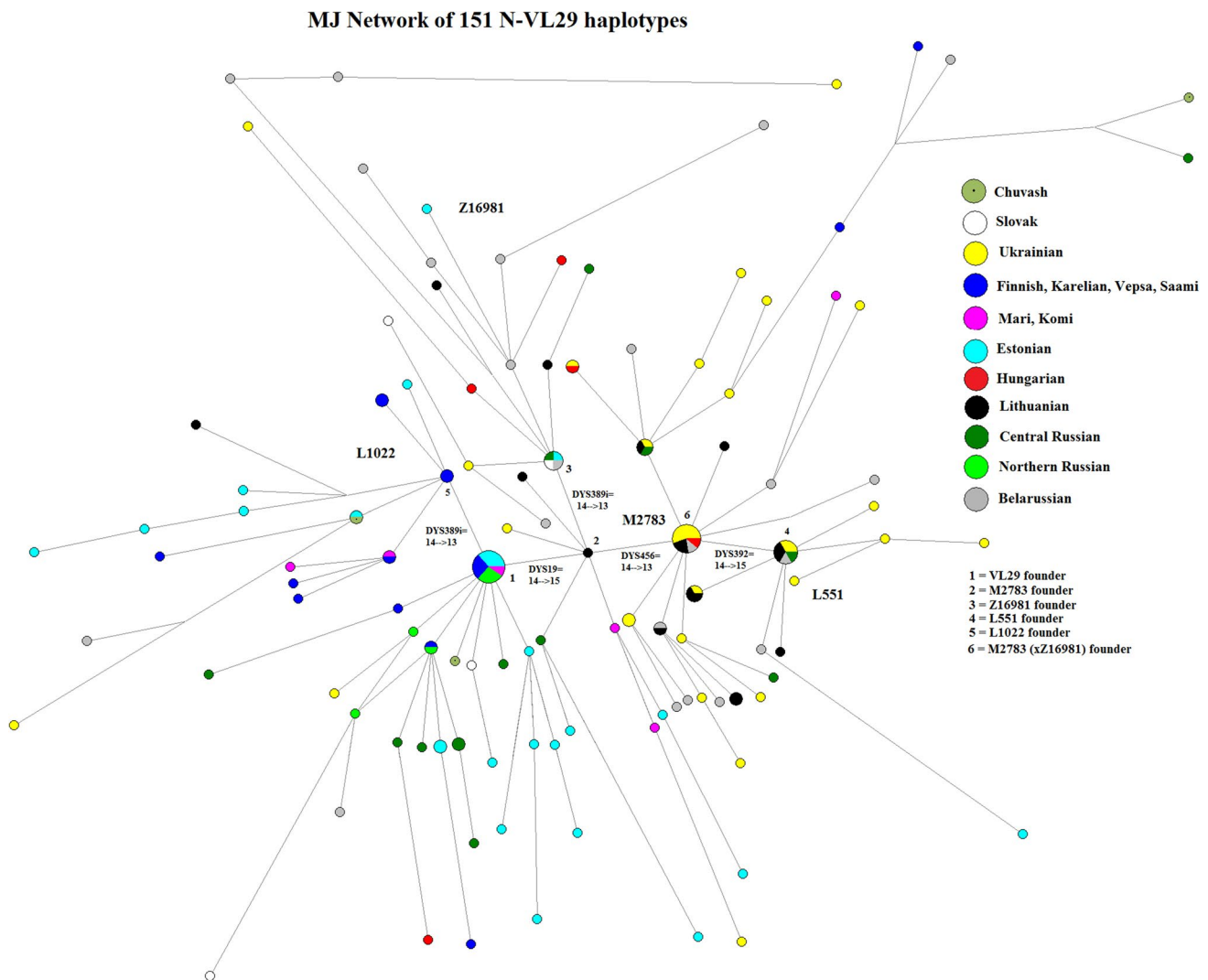


Fig. 2 Median-Joining Networks (MJ) of 151 N1c-VL29 haplotypes constructed. The *circle sizes* are proportional to the haplotype frequencies. The smallest area is equivalent to one individual

Genetic structure

Microsatellite genetic markers, such as STRs, are based on the variation in the number of repeats of DNA canonical sequences and serve as a potential tool to address population genetics and demographic history. STR loci are typically characterized by high mutation rates and high levels of polymorphisms, as well as by mutational processes that cause preferentially stepwise changes in the number of repeats and allele size. The larger the difference between two different STRs, the higher the number of mutation events is postulated to have occurred since common ancestry (Hardy et al. 2003).

Rst is an analogy of Fst based on allele size differences, so it is defined as the correlation of allele sizes between tested markers within populations. Fst is more efficient

when there are high levels of gene flow, whereas Rst better reflects population differentiation under low gene flow (Balloux and Goudet 2002). This is the reason we used Rst-based MDS analysis for population comparison.

Fst genetic distances and *p* values between 58 populations, including Bodrogköz Hungarians and Eurasian populations from published data, were calculated as shown in Online Resource 3 (ESM_3) and presented as an MDS plot (see Fig. 5). Affinities between populations may result from their common origin or from recent admixture due to geographic proximity. Genetic distances between populations, in particular, can generally be related to geographic distances, according to a model of isolation by distance (Cavalli-Sforza and Bodmer 1977). As shown in Fig. 5, the Bodrogköz Hungarian population group, along with two Hungarian,

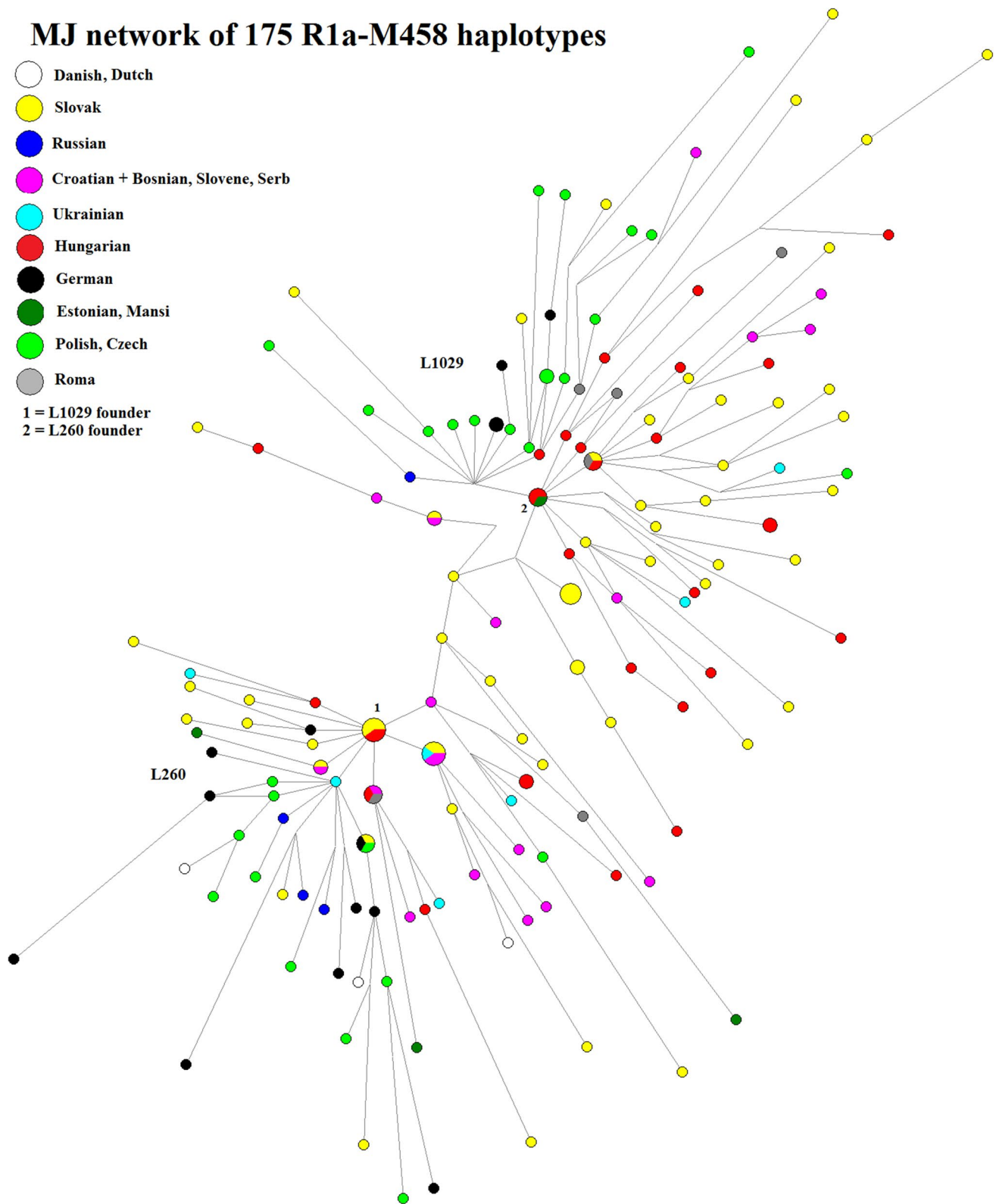


Fig. 3 Median-Joining Networks (MJ) of 175 R1a-M458 haplotypes constructed. The *circle sizes* are proportional to the haplotype frequencies. The smallest area is equivalent to one individual

MJ network of 65 J2b-M12 haplotypes

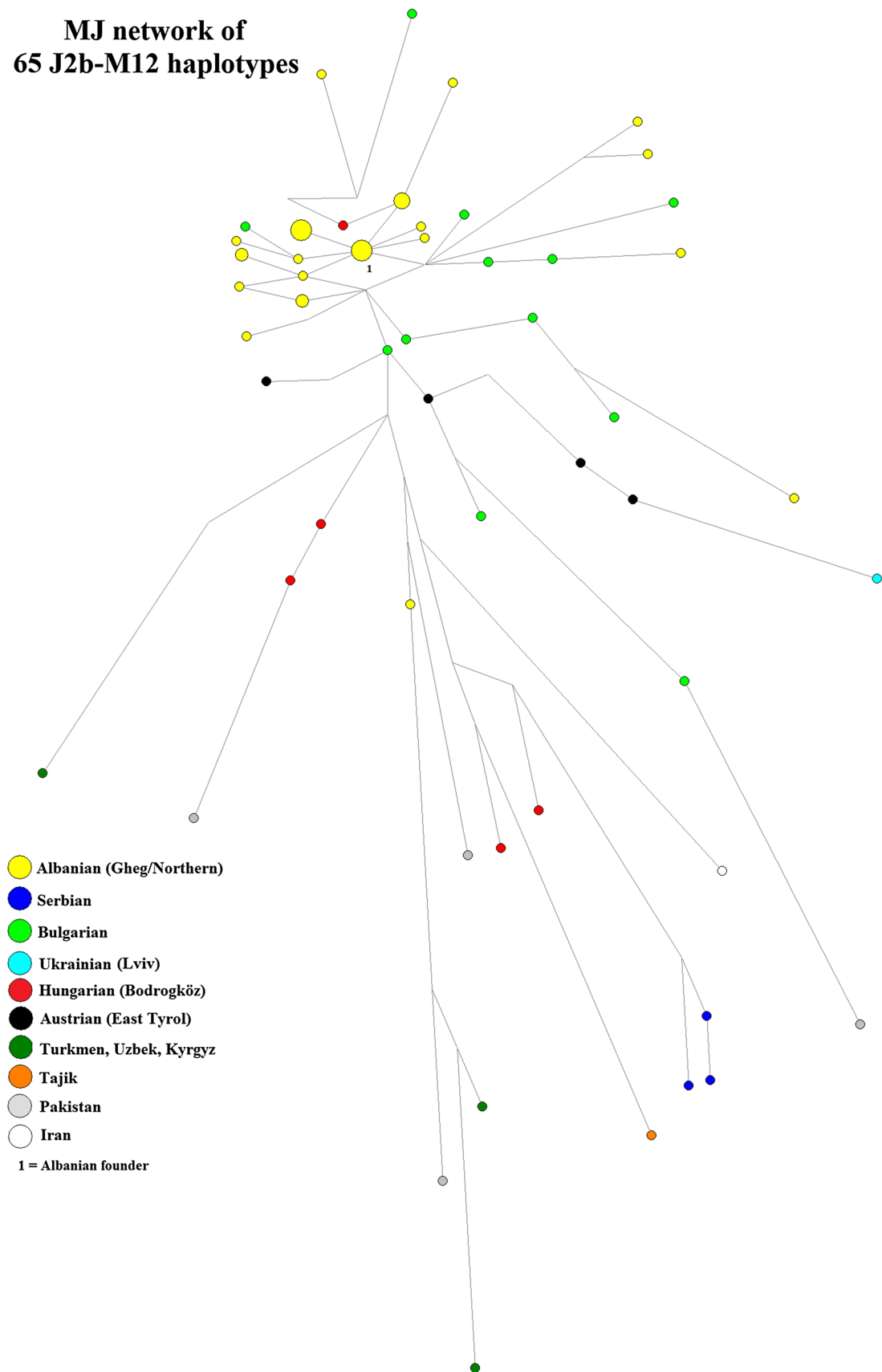


Fig. 4 Median-Joining Networks (MJ) of 65 J2b-M12 haplotypes constructed. The *circle sizes* are proportional to the haplotype frequencies. The smallest area is equivalent to one individual

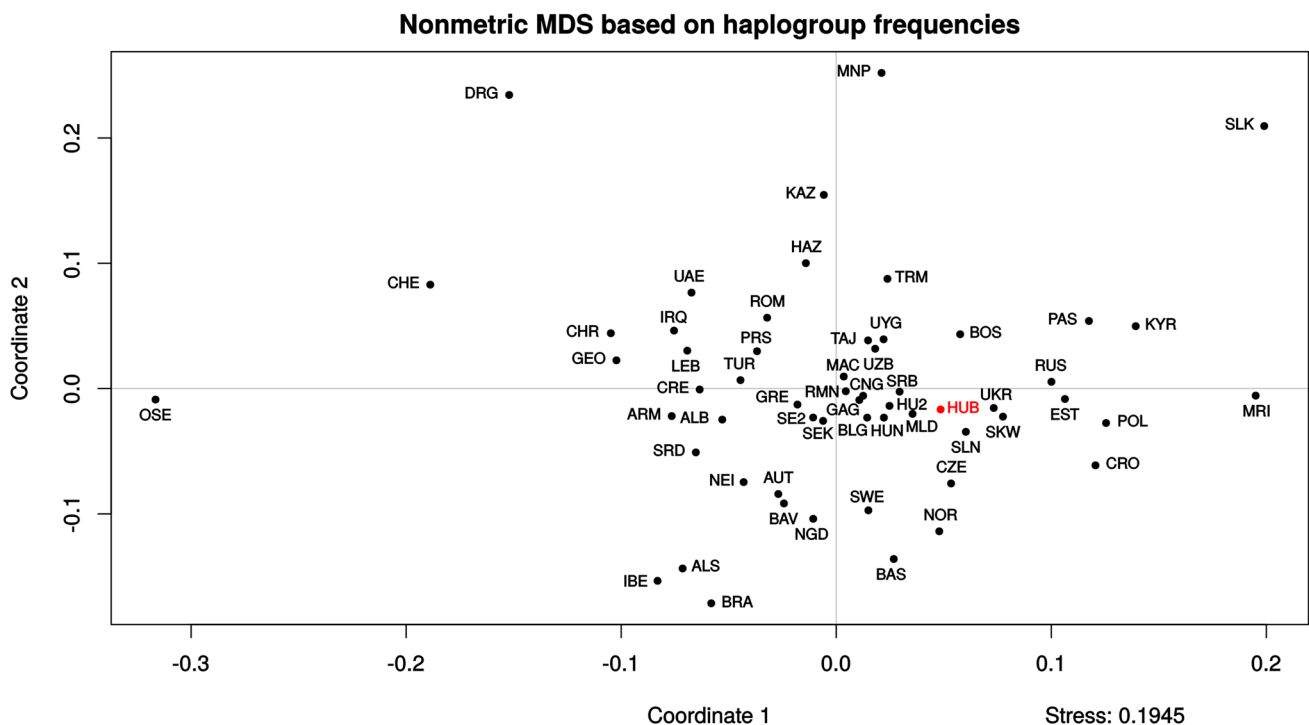


Fig. 5 Multidimensional scaling (MDS) plot constructed on F_{st} genetic distances of Y haplogroup frequencies of 58 populations compared. A three-letter coded population name is presented in Online Resources 2

Ukrainian, Western Slovakian, Slovenian, and Moldavian populations, formed a compact cluster based on close F_{st} genetic distances with non-significant p values ($p > 0.05$).

Out of the six populations, five populations were geographic neighbours of Hungarians, except for Moldavians, a finding which fits with the criteria mentioned above. These negligible genetic distances can be seen among those populations, which can be considered a single population from a genetic point of view (such pairs being Norwegians and Swedes; Bavarians, East Tyrolean Austrians and Northeast Italians; Croatians and Bosnians; Uzbeks and Tajiks in Afghanistan, and Georgians and Circassians). Apart from the East-Central European cluster, a Southeast European cluster with non-significant F_{st} genetic distances is visible among Macedonians, Bulgarians, Serbians, Romanians, and Gagauz Turks. Moldovans may form a link between the Southeast and East-Central European clusters. Interestingly, Hungarian-speaking Székely in Romania exhibited a high similarity to not only other Székely and Hungarians, but also to Greeks, who may derive from pre-Slavic genetic interactions between the two populations. Romanians are most similar to Macedonians, which could support the origin of Romanians in the Macedonian region as shown by the presence of Aromun populations in that area.

Hungarian-speaking Csángó in Romania were closest to the Gagauz sample.

We also constructed non-metric MDS based on Y-chromosomal haplotypes that consisted of 23 STR loci available from 22 populations (<http://www.yhrd.org>). As shown in Fig. 6, the Bodrozköz Hungarian population showed a non-significant R_{st} genetic distance from Ljubljana Slovenians, which is in line with the haplogroup-based F_{st} result (Slovak, Ukrainian, and Moldovan STR data were not available). Similarly, a smaller than 0.05 R_{st} distance was observed among other Hungarian, Czech, Upper Bavarian, Bosnian and Herzegovian, and Croatian and Polish populations. R_{st} genetic distances and R_{st} p values of the studied populations are presented in Online Resource 4 (ESM_4).

The genetic distances between 35 populations, including the Bodrozköz Hungarians, and other Eurasians (<http://www.yhrd.org>), were calculated based on minimum haplotypes, and were again presented as an MDS plot (Fig. 7). The Bodrozköz Hungarian population showed a close genetic affinity with non-significant R_{st} p values with surrounding populations, such as Slovenian, Ukrainian, Croatian, Slovakian, Hungarian, Ukrainian, and Romanian Csángó. R_{st} genetic distances and R_{st} p values of the studied populations are presented in Online Resource 5 (ESM_5).

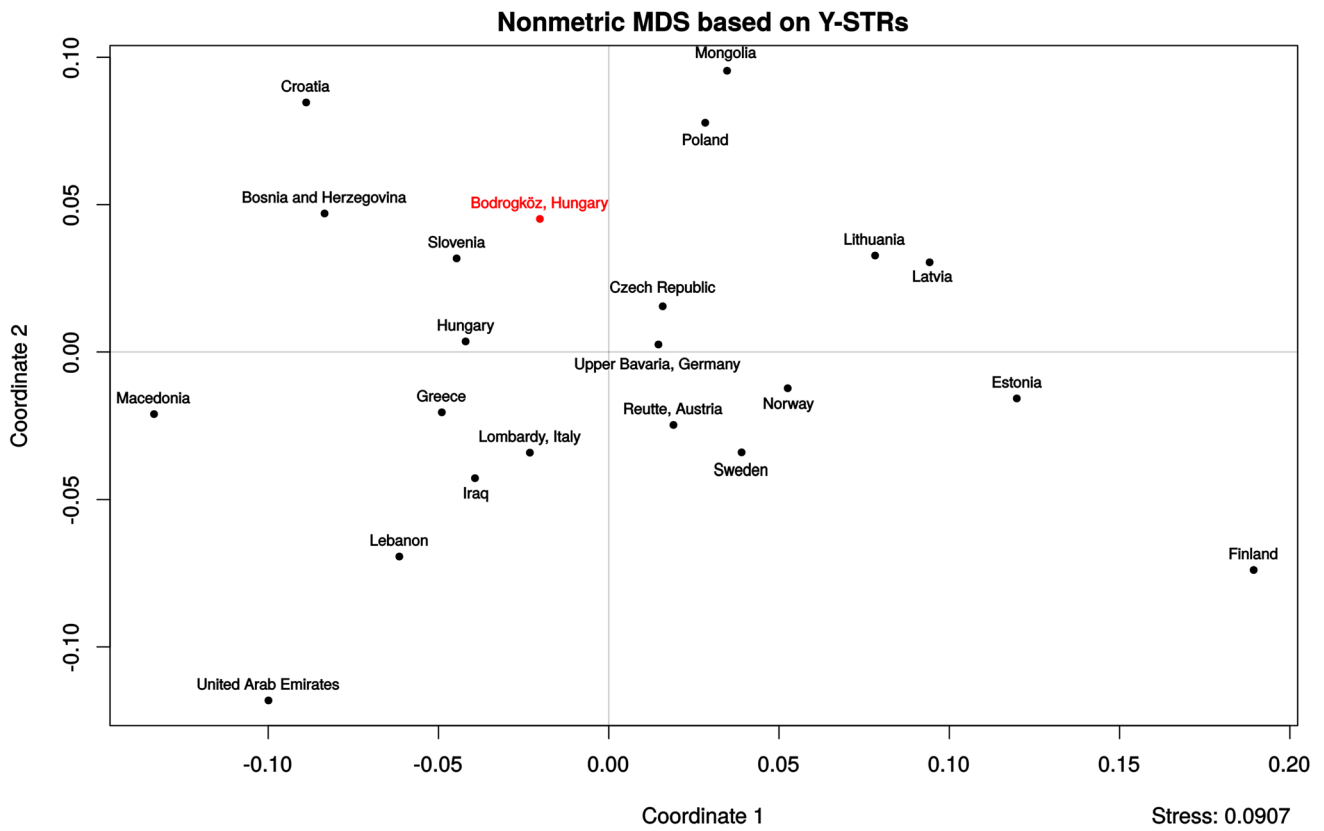


Fig. 6 Multidimensional scaling (MDS) plot constructed on Rst genetic distances of 23 STR-based Y haplotype frequencies of 22 populations compared

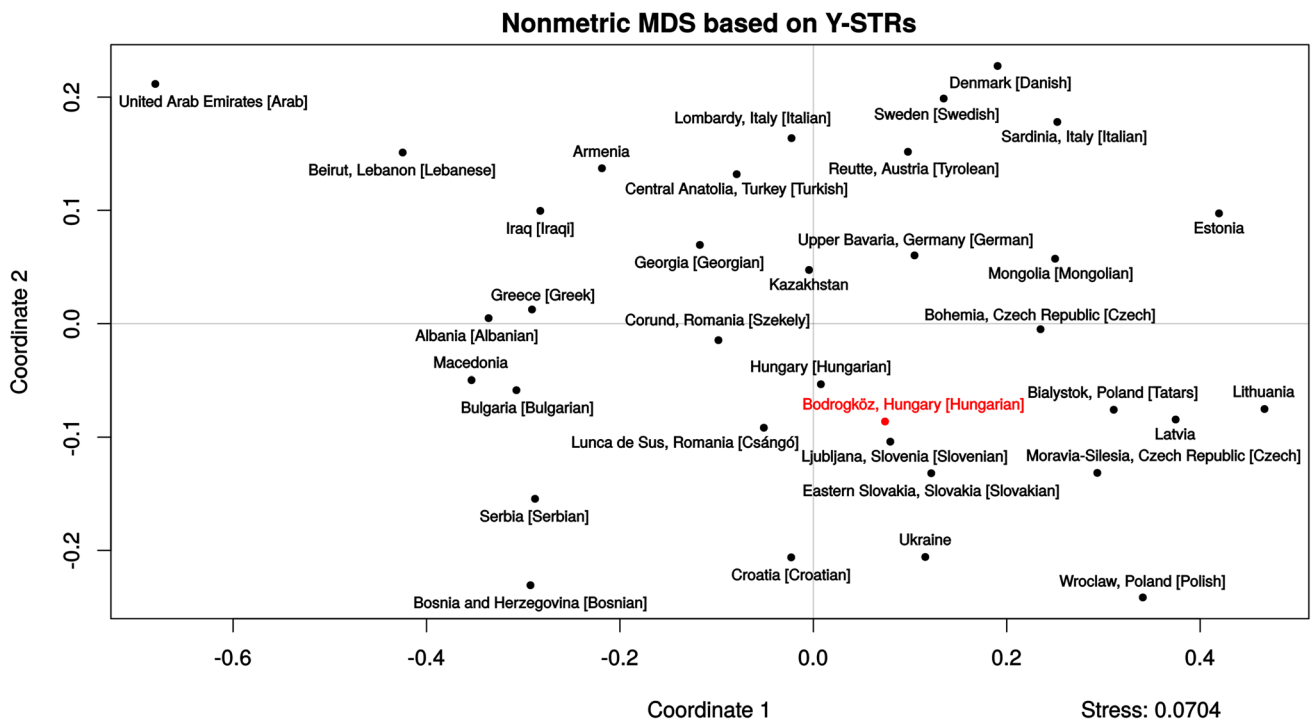


Fig. 7 Multidimensional scaling (MDS) plot constructed on Rst genetic distances of Y-chromosomal minimum haplotype frequencies of 35 populations compared

Discussion

The main objectives of this study were to estimate the level of various admixtures in the Bodrogeköz population and to determine the most likely arrival period of the analysed paternal lineages on 17 Y-STR loci comparisons with neighbouring populations. We also accounted for the pedigree mutation rate of the 17 loci (YFilter kit), as described by Goedbloed et al. (2009), considering that one mutation occurs every 668 years on average per 17 loci ($2.5 \times 10^{-3}/668$ years at 17 loci) or 705 year per 15 loci if we exclude locus DYS385. For simplicity, we defined a single mutational event (one STR molecular step) as one genetic distance (GD). This made it possible to classify the Y-haplogroups into four large groups: pre-Roman/Roman era ancient lineage (genetic distance to closest non-Hungarian match, i.e., $GD > 2$, approximately 2000 years or more), original Finno-Ugric arriving into the Carpathian Basin (closest match from Finno-Ugric-speaking population), Migration period (admixture, $GD < 2$, approximately 1300–1400 years), and post-Hungarian Conquest admixture ($GD < 2$, i.e., less than 1000 years). According to historical sources, the pre-Roman period (up to 1 year BC), Roman period (between first and fourth centuries AD), Migration period (between third and ninth centuries AD), Hungarian Conquering period (tenth century AD), and post-Hungarian Conquest period (from the 11th century AD) lasted in Carpathian Basin, Hungary (Mócsy 1987; Bóna 1987; Gyórfy 1987; Bóna 1987). These historical periods are consistent with TMRCA calculated by the pedigree mutation rate and reported in this article.

Of the analysed samples, 29.3% ($n = 43$) were of the Migration Era admixture (2 molecular steps to the closest non-Hungarian match) and another 43 samples were likely remnants of pre-Roman times (3 or more molecular steps from the closest non-Hungarian matches). This implies that on the eve of the Hungarian Conquest, half of the population of the Carpathian Basin was Roman/pre-Roman, while the other half consisted of early Slavs, Germanic peoples, and Sarmatians/Alans from the Migration Era.

We identified nine samples (6.2%) as likely Finno-Ugric lineages, with genetic matches closer to those of Uralic ethnic groups, such as Mansi and/or Estonian, than to non-Uralic groups (5 R1a-Z280, 2 R1a-M458, 1 N1c-Z1936, and 1 N1c-L1034). While haplogroup R1a-M458 has been considered “West Slavic”, our analysis showed that over half of the 30 M458 samples were not of recent Slavic admixture. For example, the TMRCA of R1a-M458, with 4700 ± 800 years based on SNPs, predates the proto-Slavic period of 300–600 AD, and this is also true of the biggest subclades: L260 2500 ± 500 YBP and L1029 2200 ± 200 YBP.

In case of R1a-Z280 and R1a-M458, deep subgrouping is needed to verify the possibility of the existence of Finno-Ugric clusters. Before this study, the predominant haplogroup of Finno-Ugric speaking peoples, namely, N1c-Tat, was not found among recent inhabitants of Hungary (Semino et al. 2000; Lahermo et al. 2000; Csányi et al. 2008). However, two out of four ancient samples from Hungarian Conquest period proved to be belonged to haplogroup N1c-Tat (Csányi et al. 2008). In our recent study, one N1c-L1034 sample was found in Bodrogeköz, which has proven to be a paternal link among Hungarians and Western Siberian Mansi (Fehér et al. 2015), and N1c-Z1936 (xL1034) was found to likely be early Ugric. Since this sample showed close genetic relationship with a sample of a warrior from a Hungarian Conquest period cemetery (our unpublished data), we could anchor the genetic descendants of a Hungarian Conqueror men. In addition, based on Y-STR analysis, one R1b-M343* sample was recent Turkic admixture from the R1b-M73 Kypchak (DYS390 = 19) cluster; however, the time of its arrival in Carpathian Basin is unknown, though it could be of Avar, Onugor, Hungarian, or Cumanian origin. Based on the Bodrogeköz population, the arriving Finno-Ugric-speaking paternal lineages likely made up no more than 10% of the population of Hungary in the tenth century AD (this does not mean that all of such lineages arrived during the Hungarian Conquest or that all Conquerors were Finno-Ugric speakers). The share of Finno-Ugric-looking paternal lineages shrank to 6.2% due to continuous influx and assimilation of neighbouring peoples in the following centuries.

The clearest post-Hungarian Conquest admixture types were the genetically Roma (2 H1a-M82) and Ashkenazi Jewish samples (J1-M267, Q-M242). Overall, according to our estimate, 35.4% of all samples ($n = 52$) were likely to be the result of post-Hungarian Conquest (896 AD) admixture in the Bodrogeköz Hungarians, meaning that they have either a zero or a single molecular step to non-Hungarian haplotypes. The admixture originated from neighbouring Germanic (G2a-L156, I1-M253, I2-M223, R1b-P312, R1b-U106) and Balto-Slavic (E-M78, I2a1-P37, N1c-VL29, R1a-M458, R1a-Z280) populations. Two samples could be identified as Albanian admixture (1 E-M78 and 1 J2b-M12).

The relatively high rate of N1c-VL29 haplotypes, as well as the close genetic similarity between the recent Moldavian, Ukrainian and Hungarian populations of Bodrogeköz, may be related to the escape of a Lithuanian prince to Hungary. Teodoras Karjotaitis, grandson of the Lithuanian Grand Prince Gediminas, had been defeated in a feud for the throne in 1393 AD, escaped to Hungary and lived under the protection of his relative, King Sigismund, in the adjacent areas of Bodrogeköz. According to historical sources, 40,000 people came to Hungary from Moldavia

and Podolia (part of present-day Ukraine), along with the Lithuanian prince (Войтович 2000).

Our study highlights the important issue that haplogroups cannot be easily linked to specific ethnic groups or admixture types without deep phylogenetic analysis and that the paternal genetic legacy of Hungarians and other neighbouring peoples is complicated. In most cases, significant migrations/demographic growth predated not only the Roman period, but also the migration waves of Germans, Slavs, and Hungarians. This is also underlined by our MDS analysis (Fig. 5), which revealed that Eastern European populations (e.g., Hungarians, Slovenians, Slovaks, Ukrainians, and Moldovans) represented close genetic affinities and a similar haplogroup composition, which, in most cases, precludes the ethnic identification of the paternal lineages.

Non-metric MDS plots constructed based on F_{st} genetic distances (Fig. 5) and R_{st} genetic distances (Fig. 6) using 23 Y-STRs showed that Bodrogeköz Hungarians, as well as two other contemporary Hungarian samples, were genetically nearly identical to Slovenians, Western Slovaks, Ukrainians and Moldovans, because they were all part of an “East-Central European” genetic cluster, regardless of their linguistic differences.

The MDS plot (Fig. 7) compares 35 population groups with Hungarian, Hungarian-speaking Székely, Csángó groups in Romania and other Eurasian populations, using minimum Y haplotypes (9 Y-STRs). The general structure of the distance matrix was depicted by an MDS plot, where the Bodrogeköz Hungarian population showed a close genetic affinity (with non-significant p values) with the surrounding populations, such as Slovenian, Ukrainian, Croatian, Slovakian, Hungarian, Ukrainian, and Romanian Csángó.

Other population groups that we compared were clearly separated from the Bodrogeköz Hungarian population. This segregation was statistically significant between most pairs of population groups. Based on R_{st} genetic distances, the Bodrogeköz Hungarians and Székely in Romania (Korond) were close to each other (0.0479) and statistically significant, suggesting close genetic relationships. Finally, we would like to highlight several Asian populations (e.g., Kazakh, Mongolian, Georgian, and Turkish in Central Anatolia) that are close to the Bodrogeköz Hungarians based on statistically significant R_{st} genetic distances, indicating an overall similarity of genetic structuring of Y-chromosome haplotypes.

In conclusion, present-day Hungarian speakers are genetically very similar to neighbouring populations, the Bodrogeköz Hungarians and to isolated Hungarian-speaking groups (Csángó and Székely in Romania). We

also conclude that they have a relatively higher presence of genetic elements that are missing from present-day neighbouring populations. However, we could classify the Y-chromosomal haplogroups into four large groups based on STR mutation events: pre-Roman/Roman ancient lineage, Finno-Ugric speakers arriving into the Carpathian Basin, Migration period admixture, and post-Hungarian Conquest admixture. Our study highlights the crucial finding that haplogroups cannot be easily linked to specific ethnic groups or admixture without deep phylogenetic analysis, as well as the fact that the paternal genetic legacy of Hungarians and other neighbouring peoples is very complicated. Our initial hypothesis that the present-day population in Bodrogeköz is likely to be more similar to the population that lived in the Conquest period, was partly confirmed. It is clear that a significantly larger database with deep haplogroup resolution, including ancient DNA data, is required to strengthen this research.

Acknowledgements This work was supported by Grant No. GF/JSZF/814/9/2015 and by encouragement from Professor Miklós Maróth; both are much appreciated. We are deeply grateful to Éva Mayer, who helped our work in every conceivable manner, and special thanks also go to Dr. Eva Susa for her support and to George L. Gabor Miklos and Ariana Gugora for English editing. We also thank all the sample donors and the laboratory assistants. Lastly, we thank Professors Gyula Viga, József Fehér and Tamás Oláh, who assisted with sample collection.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Research involving human participants and/or animals All procedures performed in studies involving human participants were in accordance with the ethical standards and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

- Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Mol Ecol* 11(4):771–783
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bíró A, Fehér T, Bárány G, Pamjav H (2015) Testing Central and Inner Asian admixture among contemporary Hungarians. *Forensic Sci Int Genet* 121–126. doi:10.1016/j.fsigen.2014.11.007 (epub 2014 Nov 15)
- Войтович ЛІ (2000) Князівські династії Східної Європи (кінець IX — початок XVI ст.): склад, суспільна і політична роль. Історико-генеалогічне дослідження. Львів, Інститут українознавства ім. І.Крип'якевича.

- Bóna I (1987) A népvándorlás kor és a korai középkor története Magyarországon. In: Bartha A (ed) Magyarország története I. Akadémia Kiadó, Budapest, pp 265–373
- Cavalli-Sforza LL, Bodmer WF (1977) The genetics of human populations. W.H. Freeman and Co., San Francisco
- Csányi B, Bogácsi-Szabó E, Tömöry G, Czibula A, Priskin K, Csőszi A, Mende B, Langó P, Csete K, Zsolnai A, Conant EK, Downes CS, Raskó I (2008) Y-chromosome analysis of ancient Hungarian and two modern Hungarian-speaking populations from the Carpathian Basin. *Ann Hum Genet* 72(Pt 4):519–534. doi:10.1111/j.1469-1809.2008.00440.x (epub 2008)
- de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am J Hum Genet* 67:1055–1061
- Di Cristofaro J, Pennarun E, Mazières S, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, Underhill PA, Vilems R, Chiaroni J (2013) Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. *PLoS One* 8(10):e76748. doi:10.1371/journal.pone.0076748 (eCollection Mar 27)
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567
- Fehér T, Németh E, Vándor A, Kornienko IV, Csáji LK, Pamjav H (2015) Y-SNP L1034: limited genetic link between Mansi and Hungarian-speaking populations. *Mol Genet Genom* 290(1):377–386. doi:10.1007/s00438-014-0925-2 (epub 2014 Sep 26)
- Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Krüger C, Roewer L, Lessig R, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M (2009) Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR Yfiler PCR amplification kit. *Int J Legal Med* 123(6):471–482. doi:10.1007/s00414-009-0342-y (epub 2009 Mar 26)
- Gyórfy Gy (1987) Magyarország története a honfoglalástól a tatárjáráig. In: Székely Gy, Bartha A (eds) Magyarország története I. Akadémia Kiadó, Budapest, pp 575–888
- Hallast P, Batini C, Zadik D, Maisano Delser P, Wetton JH, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Destro Bisoletti G, Dupuy BM, Eriksen HA, Jorde LB, King TE, Larmuseau MH, López de Munain A, López-Parra AM, Loutradis A, Milasin J, Novelletto A, Pamjav H, Sajantila A, Schempp W, Sears M, Tolun A, Tyler-Smith C, Van Geystelen A, Watkins S, Winney B, Jobling MA (2015) The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol Biol Evol* 32(3):661–673. doi:10.1093/molbev/msu327 (epub 2014 Dec 2)
- Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* 163(4):1467–1482
- Illumäe AM, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, Saag L, Agdzhoyan A, Kushniarevich A, Litvinov S, Ekomasova N, Tambets K, Metspalu E, Khusainova R, Yunusbayev B, Khusnutdinova EK, Osipova LP, Fedorova S, Utevska O, Koshel S, Balanovska E, Behar DM, Balanovsky O, Kivisild T, Underhill PA, Vilems R, Rootsi S (2016) Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. *Am J Hum Genet* 99(1):163–167
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–6123. doi:10.1016/j.ajhg.2016.05.025
- Karachanak S, Grugni V, Fornarino S, Nesheva D, Al-Zahery N, Battaglia V, Carossa V, Yordanov Y, Torroni A, Galabov AS, Toncheva D, Semino O (2013) Y-chromosome diversity in modern Bulgarians: new clues about their ancestry. *PLoS One* 8(3):e56779. doi:10.1371/journal.pone.0056779 (epub 2013 Mar 6)
- Lahermo P, Laitinen V, Sistonen P, Béres J, Karcagi V, Savontaus ML (2000) MtDNA polymorphisms in the Hungarians: comparison to three other Finno-Ugric-speaking populations. *Hereditas* 132:35–42
- Mielnik-Sikorska M, Daca P, Woźniak M, Malyarchuk BA, Bednarek J, Dobosz T, Grzybowski T (2013) Genetic data from Y chromosome STR and SNP loci in Ukrainian population. *Forensic Sci Int Genet* 7(1):200–203. doi:10.1016/j.fsigen.2012.05.007 (epub 2012 Jun 4)
- Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, Limborska SA, Herrera RJ (2009) Y-Chromosome distribution within the geo linguistic landscape of north-western Russia. *Eur J Hum Genet* 17(10):1260–1273
- Mócsy A (1987) A római kor. In: Székely Gy, Bartha A (eds) Magyarország története I. Akadémia Kiadó, Budapest, pp 199–264
- Niederstätter H, Rampl G, Erhart D, Pitterl F, Oberacher H, Neuhuber F, Hausner I, Gassner C, Schennach H, Berger B, Parson W (2012) Pasture names with romance and slavic roots facilitate dissection of Y chromosome variation in an exclusively German-speaking alpine region. *PLoS One* 7(7):e41885. doi:10.1371/journal.pone.0041885 (epub 2012 Jul 27)
- Révész L (1996) A karosi honfoglalás kori temető. Miskolc
- Sarno S, Tofanelli S, De Fanti S, Quagliariello A, Bortolini E, Ferri G, Anagnostou P, Brisighelli F, Capelli C, Tagarelli G, Sineo L, Luiselli D, Boattini A, Pettener D (2016) Shared language, diverging genetic histories: high-resolution analysis of Y-chromosome variability in Calabrian and Sicilian Arbereshe. *Eur J Hum Genet* 24(4):600–606. doi:10.1038/ejhg.2015.138 (epub 2015 Jul 1)
- Semino O, Passarini G, Quintana-Murci L, Lui A, Béres J, Czeizel A, Santachiara-Benerecetti AS (2000) MtDNA and Y chromosome polymorphisms in Hungary: inferences from the Palaeolithic, Neolithic and Uralic influences on the modern Hungarian gene pool. *Am. J. Hum Genet* 8:339–346
- Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, Di Cristofaro J, Sahakyan H, Behar DM, Kushniarevich A, Sarac J, Saric T, Rudan P, Pathak AK, Chaubey G, Grugni V, Semino O, Yepiskoposyan L, Bahmanimehr A, Farjadian S, Balanovsky O, Khusnutdinova EK, Herrera RJ, Chiaroni J, Bustamante CD, Quake SR, Kivisild T, Vilems R (2015) The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet* 23(1):124–131. doi:10.1038/ejhg.2014.50 (epub 2014 Mar 26)
- Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York. ISBN 0-387-95457-0
- Völgyi A, Zalán A, Szvetnik E, Pamjav H (2009) Hungarian population data for 11 Y-STR and 49 Y-SNP markers. *Forensic Sci Int Genet* 3(2):e27–e28. doi:10.1016/j.fsigen.2008.04.006 (epub 2008 Jun 18)
- Willuweit S, Roewer L (2015) The new Y chromosome haplotype reference database. *Forensic Sci Int Genet* 15:43–48