ANNALS OF human genetics

doi: 10.1111/j.1469-1809.2008.00440.x

Y-Chromosome Analysis of Ancient Hungarian and Two Modern Hungarian-Speaking Populations from the Carpathian Basin

B. Csányi[1,*], E. Bogácsi-Szabó[1], Gy. Tömöry[2], Á. Czibula[1], K. Priskin[1], A. Csõsz[2], B. Mende[2], P. Langó[2], K. Csete[3], A. Zsolnai[4], E. K. Conant[5], C. S. Downes[5] and I. Raskó[1]

[1] *Institute of Genetics, Biological Research Center of Hungarian Academy of Sciences, Temesvári krt. 62, H-6726, Szeged, Hungary*
[2] *Archaeological Institute of Hungarian Academy of Sciences, Úri u. 49, H-1014, Budapest, Hungary*
[3] *Department of Forensic Medicine, University of Szeged, Kossuth L. sgt. 40, H-6724 Szeged, Hungary*
[4] *Research Institute for Animal Breeding and Nutrition, Gesztenyés u. 1, H-2053 Herceghalom, Hungary*
[5] *School of Biomedical Sciences, University of Ulster, Coleraine, County Londonderry, BT52 1SA, Northern Ireland*

Summary

The Hungarian population belongs linguistically to the Finno-Ugric branch of the Uralic family. The Tat C allele is an interesting marker in the Finno-Ugric context, distributed in all the Finno-Ugric-speaking populations, except for Hungarians. This question arises whether the ancestral Hungarians, who settled in the Carpathian Basin, harbored this polymorphism or not. 100 men from modern Hungary, 97 Szeklers (a Hungarian-speaking population from Transylvania), and 4 archaeologically Hungarian bone samples from the 10th century were studied for this polymorphism. Among the modern individuals, only one Szekler carries the Tat C allele, whereas out of the four skeletal remains, two possess the allele. The latter finding, even allowing for the low sample number, appears to indicate a Siberian lineage of the invading Hungarians, which later has largely disappeared.

The two modern Hungarian-speaking populations, based on 22 Y-chromosomal binary markers, share similar components described for other Europeans, except for the presence of the haplogroup P* (xM173) in Szekler samples, which may reflect a Central Asian connection, and high frequency of haplogroup J in both Szeklers and Hungarians. MDS analysis based on haplogroup frequency values, confirms that modern Hungarian and Szekler populations are genetically closely related, and similar to populations from Central Europe and the Balkans.

Keywords: Y-chromosome haplogroups, Tat, ancient DNA, Hungarians, Szeklers

Introduction

Hungarian (Magyar) is a Finno-Ugric language, more specifically an Ugric language (Róna-Tas, 1999; Szíj, 2005), quite unrelated to the other languages of Central Europe. The Hungarian-speaking populations have a special position among the Finno-Ugrians, being the most numerous (approximately 15 million Hungarian speakers, with about 10 million in today's Hungary and another 3 million in seven neighbouring countries), and also the westernmost, surrounded entirely by Indo-European-speaking populations.

The Hungarian nation traces its history to the early Magyars (ancient Hungarians), who settled in the Carpathian Basin at the end of the 9th century after two millenia of migration from the steppe zone (Bálint, 1996). According to literature at the time of settlement the ancient Hungarians were a people who spoke a Finno-Ugric language, but had a Turkic way of life (Róna-Tas, 1995; Sindbæk, 1999; Bálint, 2006). The ancestors of the Hungarians appear to have separated from their closest linguistic relatives, the Voguls and the Ostyaks who now live in the forested areas beside the river Ob in western Siberia, around 500 BC (Fodor, 1996).

Though Hungarians linguistically belong to the Finno-Ugric language family, the ethnic components of the

* Corresponding author: Bernadett Csányi, Institute of Genetics, Biological Research Center of Hungarian Academy of Sciences, Temesvári krt. 62, H-6726, Szeged, Hungary. Phone: +36-62-599-685, Fax: +36-62-433-503, E-mail: bernadett_csanyi@yahoo.com

ancient Hungarians' population, according to recent hypothesis of historical research, could have changed significantly during the last 1000 years due to admixture with other tribes and peoples (Bálint, 2005).

The Szeklers (or Székelys) are a Hungarian ethnic group living in Transylvania, in present-day Romania. Their origin is subject to much debate, but there are two main hypotheses (Bóna, 1990; Kordé, 1994; Kristó, 2005). One is that their ancestors, after having separated from the Eskil-Bulgars of the Volga region, joined the ancient Hungarians and formed one of the tribes of the Hungarian Conquest that settled at the eastern border (Transylvania) of historical Hungary (Kristó, 2005). The second is that Szeklers are derived from the tribes that originally settled in the Danubian plain, but were removed from central Hungary to guard the Transylvanian frontier later, during the 12th century (Bóna, 1990). As a privileged people who guarded an exposed border region, Szeklers lived in comparative isolation in the mountains in Transylvania from at latest the 12th century until the 19th (Kordé, 1994).

Analyses of classical markers (Czeizel et al., 1991) have revealed that the modern Hungarians are genetically close to other European populations (Guglielmino et al., 2000). The contribution of Uralic genes for the overall Hungarian population has been estimated to be 13% (Guglielmino et al., 1990). Nevertheless, an analysis of the alleles of 24 nuclear genes by Guglielmino et al. (2000) indicated that isolated ethnic groups of Hungarians, including Szeklers, have better preserved the traces of their original gene pool than the main Hungarian population, who are very close to Slavs and Germans.

These studies were recently corrected from the field of the historical sciences (Bálint, 2006).

Studies on maternally-inherited mitochondrial markers (Lahermo et al., 2000; Bogácsi-Szabó et al., 2005; Tömöry et al., 2007) also showed that genetic variation among modern Hungarians, including Szeklers, resembles closely that found in other European populations. However, mitochondrial markers recovered from skeletal remains from the 10th and 11th centuries place the Hungarian population shortly after the invasion, on a distance matrix tree, between Asian and European populations (Tömöry et al., 2007).

These investigations, then, have consistently demonstrated that the bulk of the modern Hungarian population is genetically indistinct from the surrounding Indo-European-speaking populations; but maternal lineages in early Hungarians are more distinct.

Paternal lineages, which usually give a higher geographical resolution than maternal, have been studied for 11 biallelic Y-chromosomal markers in a small sample of modern Hungarians, and again appear similar to the surrounding populations (Rosser et al., 2000).

One specific Y-chromosomal base substitution (Tat, T→C) a relatively recent event (95% confidence interval 3140–6200 years, Lahermo et al., 1999) is a valuable marker in Finno-Ugric population studies (e.g. Zerjal et al., 1997). According to published data the C allele of the Tat polymorphism is widespread in all Uralic-speaking populations studied so far, except that it is absent or extremely rare among modern Hungarian-speaking populations (Zerjal et al., 1997; Lahermo et al., 1999; Rootsi et al., 2000; Semino et al., 2000a, 2000b; Tambets et al., 2001, 2004). This rarity could be due to extensive gene flow into the Hungarian-speaking population before their arrival in modern Hungary, or in the eventful centuries afterwards.

To clarify this, in the present study, we therefore screened for the Tat polymorphism in ancient DNA from skeletal remains from the age of the Hungarian Conquest.

In addition, to gain further insight into the paternal genetic diversity of the modern Hungarian-speaking populations, we have typed additional markers from the non-recombining region of the human Y chromosome (NRY); we present the analysis of 22 NRY biallelic polymorphisms in Hungarian samples, predominantly from the Great Hungarian Plain, and Szeklers from Transylvania. The results are compared with data from other European populations studied by Semino et al. (2000a), and the phylogeographic context of the Y chromosome pool of the populations studied has been analysed.

Materials and Methods

Samples

In this study 100 Hungarian and 97 Hungarian-speaking Szekler individuals were involved. All the Hungarian paternally unrelated, healthy males selected for the analysis had a birthplace in different parts of Hungary, but the most represented area (90 samples) was the Great Hungarian Plain (occupying the southern and eastern part of Hungary). The Szekler sample consists of 97 unrelated, healthy Szekler volunteer donors born and living in Corund, Transylvania, Romania (Figure 1). Most of the DNA samples (94 Hungarian and 97 Szekler samples), extracted from blood, were anonymously obtained from the DNA collection of the Department of Forensic Medicine (University of Szeged, Hungary). Six additional Hungarian DNA samples, isolated from hairbulbs, were from our DNA depository. The Szekler samples are identical to those recently studied for Y-STR variation (Beer et al., 2004).

For ancient DNA, bones from 8 skeletons were studied, but in only four cases was DNA successfully isolated. These four samples were derived from three different well-documented Hungarian excavations from 10th century cemeteries (Figure 1). Burial sites and bones were archeologically and anthropomorphically well defined before analysis (Table 1). Three ancient remains (anc4, anc21, anc28) were excavated with rich grave goods,

Figure 1 Map showing the locations of the studied samples: Numbers 1, 2, 3 refer to locations of the cemeteries where the bone samples were excavated. The Great Hungarian Plain includes the birthplaces of most of the Hungarian males studied (90 out of 100 samples).The Szekler males studied live in Transylvania, Romania.

unambiguously typical of funeral practices of Hungarian conquerors (Szõke, 1962; Révész, 1996; Mesterházy, 1997). These samples were from two burial sites, in the same county (Békés) in the eastern part of the Great Hungarian Plain. The fourth sample (anc19) derived from a cemetery located in the western part of Hungary, with poor grave goods, possibly the burial of a soldier. These ancient samples are a subset of those 27 previously typed for mitochondrial DNA in our laboratory (Tömöry et al., 2007). The denomination of the samples is identical to that used in that study. Gender of the remains was initially determined by anthropometric evaluations.

DNA Extraction

DNA from blood samples was extracted using the salting-out procedure (Miller et al., 1988) or Chelex-based method (Walsh et al., 1991) and from the hair samples by the Chelex method.

DNA from the archaeological remains was extracted from femoral bones according to the method of Kalmár et al. (2000). The DNeasy Tissue Kit (Qiagen, Hilden, Germany) was also used, following the manufacturer's recommendations. A modified method, which combines the two protocols, was also applied (Tömöry et al., 2007).

Whole Genome Amplification

After successful and contamination-free mtDNA amplification from the archaeological samples, Y-chromosomal DNA recovery was attempted. In order to enhance the efficiency random primers and the mIPEP method (Hanson & Ballantyne, 2005) was used. The mIPEP amplification was performed with 5–10 μl bone DNA extract in a 25 μl reaction volume consisting of: 0.8 mM dNTPs (Fermentas, Burlington, ON, Canada), 2.5 mM $MgCl_2$ (Fermentas), 1x High Fidelity PCR Buffer

Table 1 Sample Characteristics

		Anthropological features			
Sample[a]	Site[b]	Sex and Age	Taxonomy	Estimated Age	Type of grave goods
anc4	Szabadkígyós-Pálliget (2)	Male, 25–30 years	Europid (Cromagnoid)	middle 10th century	classical
anc19	Mözs-Szárazdomb (3)	Male, 24–28 years	–	middle 10th century	poorer
anc21	Örménykút (1)	Male, 61–67 years	Europo-Mongolid	late 10th century	classical
anc28	Szabadkígyós-Pálliget (2)	Male, 45–50 years	–	middle 10th century	classical

[a]Denomination of the samples is identical to that used in our previous study (Tömöry et al., 2007).
[b]Numbers in brackets refer to cemeteries on the map in Figure 1.

(Fermentas), 10.5 U High Fidelity Enzyme Mix (Fermentas), 100 μg/ml BSA (New England Biolabs, Ipswich, MA, USA) and 40 μM PEP primer (5′-NNNNNNNNNNNNNNN-3′). The amplification protocol was 50 cycles of 94°C for 1 min, 37°C for 2 min, and 0.1°C/s ramp to 55°C for 4 min. The High Fidelity PCR Enzyme Mix contains a highly processive Taq polymerase and a second Pfu DNA polymerase that exhibits a 3′-5′ exonuclease activity. 4–10 microliter aliquots of the mIPEP amplification product were used in subsequent Y-chromosomal analyses.

Analysis of the Tat Polymorphism in Ancient Samples

Screening the Tat polymorphism was carried out by amplifying a 112 bp fragment using the primers Tat1 and Tat3 (Zerjal et al., 1997), and a subsequent digestion with HpyCH4 IV enzyme (New England Biolabs). The reverse primer was 5′-labeled with Cy5 dye. PCR reactions were performed in a volume of 25μl containing 200 μM of each of the dNTPs (Fermentas), 1 μM of each primer, 3 mM $MgCl_2$ (Applied Biosystems, Foster City, CA, USA), 1 U AmpliTaq Gold DNA Polymerase (Applied Biosystems), 1 x PCR Buffer, 100 μg/ml BSA (New England Biolabs) and 5–10 μl aliquots of the mIPEP amplification product. The standard amplification protocol was: 94°C for 6 min, followed by 38 cycles of 30 sec at 93°C, 1 min at 56°C, 45 sec at 72°C, and a final extension at 72°C for 5 min. PCR products and restriction fragments were resolved and detected on 4% denaturing polyacrylamide gel (ALFexpress II DNA Analyser System, Amersham Biosciences, Piscataway, NJ, USA). Fragment lengths were determined by ALFwin Fragment Analyser 1.00 software. For accurate typing an external ladder (ALFexpress Sizer 50–500 bp, a ladder of labeled fragments with increments of 50 bp from 50 bp to 500 bp), and internal molecular weight standards (ALFexpress Sizer 100 bp and 150 bp) were used (Figure 2). Control DNA samples for the ancestral and its derivative mutant state were always applied simultaneously with experimental samples.

Contamination Prevention

To prevent any possible risk of contamination during the handling and analysis of the ancient remains strict precautions were taken (Bogácsi-Szabó et al., 2005; Tömöry et al., 2007). Controls were routinely used: an extraction control to check the purity of the mock DNA extraction with no bone added and an amplification control to check the purity of the PCR reagents and set-up with no DNA added to the PCR reaction. All male investigators (excavator, anthropologist) handling the material were typed for Tat polymorphism.

To authenticate the results at least two independent experiments were carried out by female researchers. In case of each ancient sample, all steps of work (sample preparation, bone powdering, DNA extraction, amplification, post-PCR analysis) were repeated at least twice, to obtain consistent results. Sample preparations, bone powdering, extractions and setting up of PCR reactions were carried out in laboratories dedicated to ancient DNA work. Each bone sample was powdered by two researchers in independent laboratories (Institute of Genetics, Biological Research Center of HAS (BRC lab) and Archaeological Institute of HAS), where from each ancient human remain, the different part of the femoral diaphysis was removed. The further steps of analysis (DNA extraction, amplification and post-PCR analysis) were performed at the BRC lab in separate places by two different female scientists. The Tat T/C genotype was considered reliable when results were reproduced from at least two independent DNA extractions and amplifications.

Y-Chromosome Analysis of Modern DNA Samples

Twenty-two biallelic markers (M96, M89, M9, M45, M35, M78, M170, M253, P37, M26, M201, P15, M304, M267, M172, M102, M67, M92, Tat, M173, M17, M269) were examined in hierarchical order in agreement with the Y-chromosome phylogeny (Y Chromosome Consortium, 2002; Jobling & Tyler-Smith, 2003). For most of the markers, new primers were designed to amplify shorter fragments (Table 2, Table 3).

All samples were surveyed for the M96, M89, M9 and M45 markers. These markers were typed by multiplex PCR followed by SNaPshot technique (see below and Table 2).

Additional genotyping of samples was restricted to markers on the appropriate branch of the YCC haplogroup tree (Jobling & Tyler-Smith, 2003). Markers M35, M78, M170, M253, P37, M26, M201, P15, M304, M267, M172, M102, M67, M92, Tat, M173, M17 and M269 were analysed by the PCR-RFLP assay.

In some cases the dCAPS (derived cleaved amplified polymorphic sequence) method was used to introduce a restriction enzyme recognition site, which includes the SNP, into the PCR product by the forward or reverse primer containing a mismatch to the template DNA (Neff et al., 2002) (Table 3). Control DNA samples for the ancestral and mutant states of each binary marker were always included.

The YCC updated phylogeny (Jobling & Tyler-Smith, 2003) was followed to assign and determine the haplogroups.

Multiplex PCR (Markers M96, M89, M9, M45)

A modified version of the protocol described by Belgrader et al. (1996) and Paracchini et al. (2002) was used. In this procedure the amplification takes place in two stages.

The first reaction was performed in a volume of 12.5 μl containing 2.5 μl blood or hair DNA extract, 1x AmpliTaq Gold Buffer (Applied Biosystems), 4 mM $MgCl_2$, 400 μM of each of the dNTPs, 0.32 μM of each PCR hybrid primer (Table 2), 1.5 U AmpliTaq Gold DNA Polymerase (Applied Biosystems). The reaction consisted of denaturation at 94 °C for 9 min, followed by 15 cycles of 94 °C for 30 sec, 63 °C for 30 sec and 72 °C 1 min, with a final extension at 72 °C for 3 min. An equal volume of complete PCR buffer containing 1 μM zip code primers, 4 mM $MgCl_2$ and 1.5 U enzyme was added and cycled for an additional 30 rounds as above, except that the annealing temperature was 59°C. The

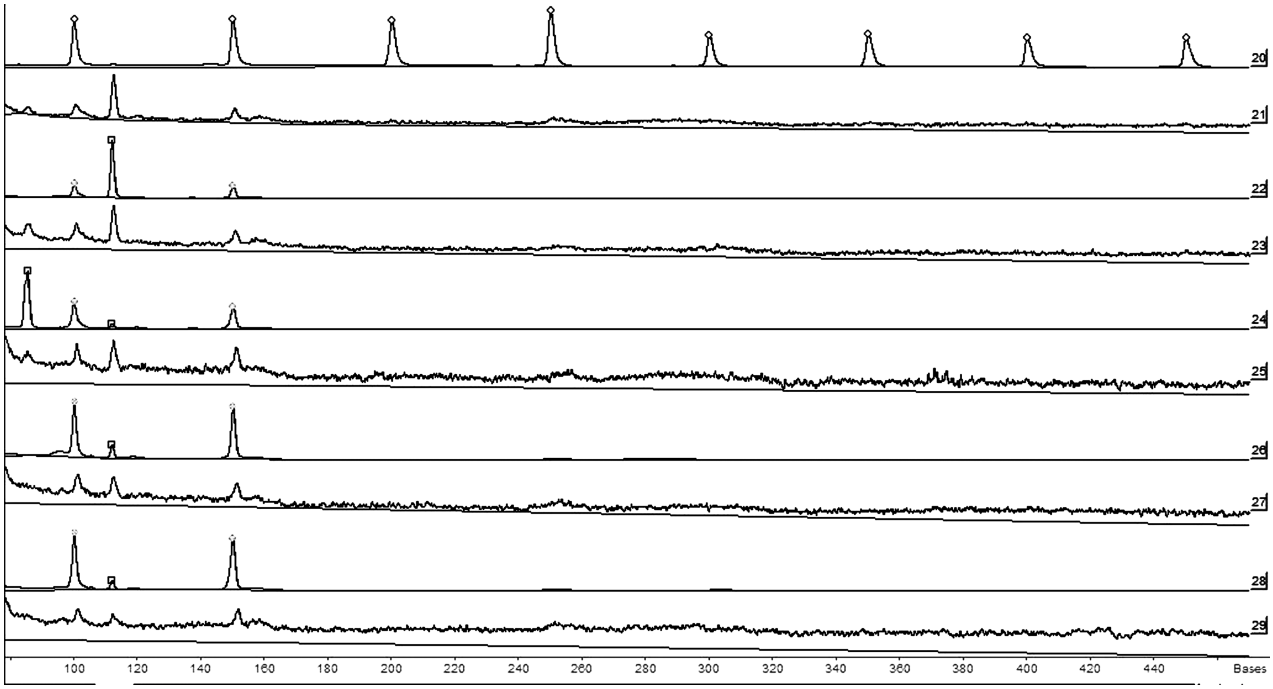


Figure 2 Electrophoretogram displaying PCR-RFLP analysis of Y-chromosomal Tat polymorphism on ancient DNA from two samples from the 10th century. Peaks represent fluorescent intensities of Cy5 dye-labeled DNA products. The reverse primer was fluorescently labeled with dye Cy5. The size of the products, in bases, is shown along the x axis.
lane 20: ALFexpress Sizer 50–500 bp (from 100 bp to 450bp)
lane 22, lane 26: the 112 bp long amplified products in case of anc28 and anc19 bone samples, respectively, with 100 and 150 bp long internal markers
lane 24, lane 28: the results of restriction cleavage with HpyCH4 IV enzyme; the presence of 85 bp long fragment in case of bone sample anc28 (lane 24) shows that this sample carries the mutant C allele, while its absence in lane 28 indicates the ancestral T allele on Y chromosome of the sample anc19.
No samples were loaded into lanes 21, 23, 25, 27, 29, to prevent contamination of samples from neighbouring lanes.

hybrid primers have locus-specific 3′-ends and 5′-ends corresponding to zip code primers.

Four microliters of PCR product was run on 8% native polyacrylamide gel and visualised after ethidium bromide staining by ultraviolet transillumination. The size of the products was determined with a GelBase gel documentation system.

Negative controls were used, in which female DNA and no template was added to the PCR mix, respectively. As a positive control previously sequenced male DNA was applied.

Genotyping

Following the multiplex PCR amplification, unconsumed PCR primers and dNTPs were removed by adding 1.2 μl ExoSap-IT (USB) reagent, consisting of Exonuclease I and Shrimp Alkaline Phosphatase, to 3 μl PCR products for 15 min at 37 °C, followed by the enzyme inactivation at 80 °C 15 min. Three microliters of the purified product was submitted to primer extension reaction, using the ABI PRISM SNaPshot ddNTP Primer Extension Kit (Applied Biosystems) according to the instructions of the manufacturer.

Concentrations of the minisequencing primers (Table 2) were 0.1 μM, 0.6 μM, 0.1 μM, 0.3 μM for M96, M89, M9, M45 markers, respectively. After Shrimp Alkaline Phosphatase treatment (USB) 0.25 μl of the fluorescently labeled dideoxy-terminated fragments were analysed on ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Genotype assignment was accomplished by Genotyper Software Version 3.7 (Applied Biosystems). Following the manufacturer's recommendations, the difference in minisequencing primer lengths was 5 nucleotides, since the migration distance does not necessarily refer to the length of the one-base elongated primer. Their lengths were adjusted with polyT or in case of the M96 marker with 'taat' to avoid secondary structures (Table 2).

This standard procedure (Multiplex PCR and Genotyping) was successfully employed for the modern samples and partly on degraded DNA also, namely in the case of anc19, where DNA fragments, which carry the M45 and M9 markers were amplified. The strategy was the same as described above, except that the mIPEP amplification product (4 μl) was used as template and ten microliters of PCR product was run on polyacrylamide gel after the multiplex PCR.

Table 2 Primers for PCR/SNaPshot typing of Y-Chromosomal binary markers M96, M89, M9 and M45

	Primers			
	PCR		SNaPshot Reaction	
Marker/Mutation	Sequence*	Amplicon size[a]	Sequence	Length (mer)
M96 G→C	5′-ggagcacgctatcccgttagac**GTTGCCCTCTCACAGAGCAC-**3′ 5′-cgctgccaactaccgcacatg**GAAGAGATTCACCCACCCAC-**3′	200 bp (157 bp)	5′-taatGAAAACAGGTCTCTCATAATA-3′	25
M89 C→T	5′-ggagcacgctatcccgttagac**TCCTATGAGGTGCCATGAAA**-3′ 5′-cgctgccaactaccgcacatg**GGATCACCAGCAAAGGTAGC**-3′	180 bp (137 bp)	5′-ttttttttttttttttCTCAGGCAAAGTGAGAGAT-3′	35
M9 C→G	5′-ggagcacgctatcccgttagac**TGCAAAGAAACGGCCTAAG**-3′ 5′-cgctgccaactaccgcacatg**GCATAATGAAGTAAGCGCTACC**-3′	164 bp (121 bp)	5′-ACGGCCTAAGATGGTTGAAT-3′	20
M45 G→A	5′-ggagcacgctatcccgttagac**GTGGACTTTACGAACCAACCT-**3′ 5′-cgctgccaactaccgcacatg**CCTGGACCTCAGAAGGAGCT-**3′	149 bp (106 bp)	5′-ttttttttttCCTCAGAAGGAGCTTTTTGC-3′	30

*Genomic sequence in bold, zip sequence in lower case.

[a]Numbers in parentheses indicate the size of the specific PCR products.

The zip code primer sequences were published by Belgrader et al. (1996).

The locus-specific part of PCR primers for markers M9, M45 and the reverse primer of marker M96 were designed in our study to get shorter fragments.

The PCR primers for marker M89 were published by Paracchini et al. (2002) and the forward primer of marker M96 was published by Underhill et al. (2000).

Table 3 Primers and Parameters for PCR/RFLP Typing of Y-Chromosomal Binary Markers M35, M78, M170, M253, P37, M26, M201, P15, M304, M267, M172, M102, M67, M92, Tat, M173, M17 and M269

Markers			**Primers**			**RFLP**		
Name	Mut.	Reference	Sequence[a]	Reference	Annealing temperature	Enzyme	Ancestral (bp)	Derived (bp)
M35	G > C	Underhill et al., 2000	5′-gagagggcatggtccctttc	this study	57 °C	Bts I	88	60 + 28
			5′-tttcggagtctctgcctgtg	//				
M78	C > T	Underhill et al., 2000	5′-tcgacatgaacacaaattgatacactt	Onofri et al., 2006	58 °C	Aci I	45 + 46	91
			5′-aaagcaagtactatgaccagc	this study				
M170	A > C	Underhill et al., 2000	5′-catattctgtgcattatacaaat	this study	60 °C	Nla III	57 + 20	77
			5′-acacaacccacactgaaaaa**c**a	//				
M253	C > T	Cinnioğlu et al., 2004	5′-tccatttagctgatctgtttc	this study	58° C	Hinc II	117 + 34	151
			5′-caaggactcattcaatgaaga	//				
P37	T > C	Karafet et al., 2001	5′-gcatagtgatagggtgggattgg	this study	67° C	HpyCH4 III	95	29 + 66
			5′-cctggcaaaggcgggaatcc	//				
M26	G > A	Underhill et al., 2000	5′-gacagctggcttaccagtggt	this study	59° C	Aci I	81 + 29	110
			5′-gaatttcataggccattcagtgttctc**c**g	//				
M201	G > T	Underhill et al., 2001	5′-tatgcatttgttgagtatatgtcaaat	Onofri et al., 2006	55° C	Bbs I	129 + 34	163
			5′-tccaacactaagtacctattacgaa**g**a	this study				
P15	C > T	Hammer et al., 2000	5′-cctcacatgaatagagccaa	this study	52° C	Mlu I	41 + 21	62
			5′-actttcatctgccttca**c**gc	//				
M304	A > C	Cinnioğlu et al., 2004	5′-tgtaacaaacagtatgtggga	this study	56° C	NmuC I	81	54 + 27
			5′-ataccaaaatatcaccagttg	//				
M267	T > G	Cinnioğlu et al., 2004	5′-tataccaagtctggatagcgga	this study	50° C	Mnl I	71	41 + 30
			5′-cttccacacaaaatactgaac**c**t	//				
M172	T > G	Underhill et al., 2000	5′-ctgcctctcagtatcaacagt	this study	58° C	Hinf I	91	68 + 23
			5′-taataattgaagacctttt**g**ag	//				
M102	G > C	Underhill et al., 2000	5′-agataaaattcacatagtgca	this study	54° C	NmuC I	106	24 + 82
			5′-tccttaatctctaggggt	//				
M67	A > T	Underhill et al., 2000	5′-aaaggctctatcttcaagtacgtgtccta	Onofri et al., 2006	59° C	Nde I	130	104 + 26
			5′-cacttgttcgtggacccctcta**c**at	this study				
M92	T > C	Underhill et al., 2000	5′-ttggacttaaaggtggcttg	this study	54° C	HpyCH4 IV	161	30 + 131
			5′-ttcagaaactggttttgtgtcc	Underhill et al., 2000				
Tat	T > C	Zerjal et al., 1997	5′-gactctgagtgtagacttgtga	Zerjal et al., 1997	56° C	HpyCH4 IV	112	27 + 85
			5′-gaaggtgccgtaaaagtgtgaa	//				
M173	A > C	Underhill et al., 2000	5′-ttcttacaattcaagggcattt**g**gaac	this study	67° C	Nla IV	98	25 + 73
			5′-gctgcagttttcccagatcct	Onofri et al., 2006				
M17	G > -	Underhill et al., 1997	5′-agagtttgtggttgctggttgttacg**c**g	this study	56° C	Mlu I	117	24 + 92
			5′-tgatgtagagacatctgaaacccac	//				
M269	T > C	Cruciani et al., 2002	5′-ggggaatgatcagggtttgg	this study	57° C	ScrF I	88	27 + 61
			5′-acttcttttgtgtgccttctgag	//				

[a]The mismatching nucleotide is underlined and boldfaced.

PCR-RFLP Analysis

For markers M35, M78, M170, M253, P37, M26, M201, P15, M304, M267, M172, M102, M67, M92, Tat, M173, M17 and M269 the PCR amplifications were performed in reaction volumes of 25 μl with 1 U AmpliTaq Gold DNA Polymerase (Applied Biosystems), 1 x PCR Buffer, 3 mM $MgCl_2$, 200 μM of each of the dNTPs, 1–1 μM primers and 5 μl DNA template. The PCR cycles were: initial denaturation of 6 min at 94°C, followed by 33 cycles of 20 sec at 93°C, 20 sec at the primer-specific annealing temperature (50–67°C)(listed in Table 3), 30 sec at 72°C and final extension of 5 min at 72°C. To increase the specificity of the PCR reaction in the case of markers M170 and M267, prior to the 33 cycles, 10 cycles of 20 sec at 93°C, 20 sec at higher annealing temperature (64°C-M170 or 54°C-M267) and 30 sec at 72°C was needed and in the case of marker P37 we used 0.5 μM primers and 3 μl DNA template.

Following the restriction cleavage (Table 3) genotyping was performed after the separation of the restriction products in 8% polyacrylamide gel and visualisation with ethidium bromide staining.

Statistical Analyses

Population genetic diversity measured as H (Nei, 1987), analysis of molecular variance (AMOVA) and pairwise Fst genetic distances (Reynolds et al., 1983) based on haplogroup frequencies were calculated by using the Arlequin package, version 2.000 (Schneider et al., 2000). The statistical significance of Fst values was estimated by permutation analysis using 10 000 permutations.

Multidimensional scaling (MDS) analysis of pairwise Fst values was performed using the SPSS package version 5.0 (SPSS, Inc., Chicago, IL, USA).

Results

Analysis of Tat Polymorphism

The analysis of the Tat polymorphism in 100 Hungarian, 97 Szekler and 4 ancient samples through PCR amplification and restriction cleavage revealed that, among the modern population, only one Szekler man carries the mutant C allele; whereas out of the four successfully typed 10^{th}-century skeletal remains two (anc21 and anc28), possess the mutation, while the samples anc4 and anc19 carry the ancestral T allele. Figure 2 shows a representative electrophoretogram of ancient DNA from samples anc28 and anc19.

Figure 3 denotes the multiplex amplification (markers M9, M45) of ancient DNA from sample anc19. The presence of the ancestral C allele of the M9 polymorphism reinforced the fact that this sample does not harbour the mutant Tat C allele, in accordance with standard Y-chromosome phylogeny. Out of four Y-chromosomal DNA fragments, containing the four basal markers (M96, M89, M9 and M45), the two shorter fragments (106 bp and 121 bp long specific PCR products) were successfully amplified, where the signal reflecting the M45 marker was stronger. Such an inverse relation between the efficiency of amplification and the length of the amplified product is typical of DNA retrieved from archaeological remains and results from damage and degradation of the DNA (Handt et al., 1994). The loci M89 and M96 could not be amplified, probably because they are in the higher molecular weight range.

No one from the Archaeological Institute who has handled the bones (excavator, anthropologist) has the Tat C

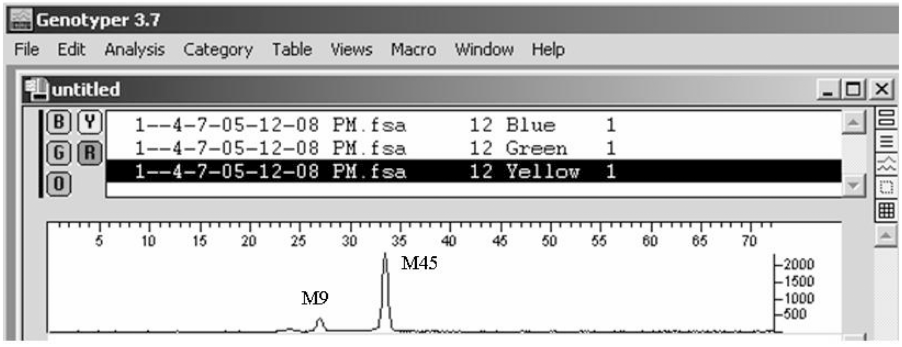
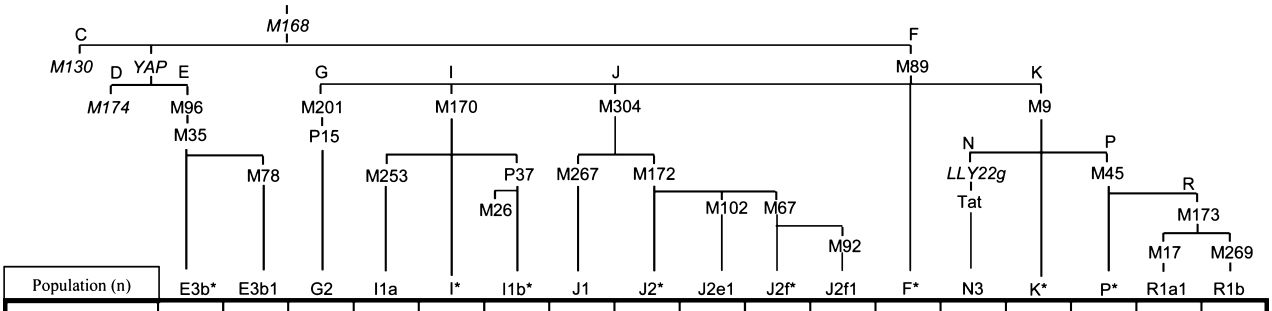


Figure 3 Electrophoretic pattern of markers M9 and M45 genotypes in case of sample anc19. Out of four Y-chromosomal DNA fragments, containing the four basal markers (M96, M89, M9, M45), two shorter fragments (106 bp and 121 bp long specific PCR products) were successfully amplified.
In case of these markers the ancestral C allele was typed. The ancestral state at marker M9 indicates that this sample could not carry the derived Tat C allele in accordance with Y-chromosome phylogeny (YCC, 2002; Jobling & Tyler-Smith, 2003). This finding confirmes the result of the PCR-RFLP analysis of this sample.



Population (n)	E3b*	E3b1	G2	I1a	I*	I1b*	J1	J2*	J2e1	J2f*	J2f1	F*	N3	K*	P*	R1a1	R1b
Hungarians (100)	1.0	9.0	3.0	8.0	3.0	13.0	3.0	8.0	4.0	1.0	0.0	1.0	0.0	1.0	0.0	30.0	15.0
Szeklers (97)	1.0	8.2	5.2	16.5	0.0	5.2	10.3	3.1	7.2	0.0	1.0	0.0	1.0	0.0	3.1	18.6	19.6

Figure 4 Phylogeny of Y-chromosome haplogroups studied and their frequencies in Hungarian and Szekler populations. Italicized markers were not typed, but their allelic states can be deduced from the YCC 2003 phylogenetic tree (Jobling & Tyler-Smith, 2003).

mutation, which strongly supports the authenticity of the ancient DNA.

Y-Chromosome DNA Analysis of Modern Samples

In 197 individuals 22 biallelic markers (M96, M89, M9, M45, M35, M78, M201, P15, M170, M253, P37, M26, M304 (M304 is equivalent to 12f2 (Cinnioğlu et al., 2004), M267, M172, M102, M67, M92, Tat, M173, M17, M269) were examined. After PCR amplification the SNaPshot method and restriction cleavage were used.

The markers used allowed the classification of these samples into the E, F*, G, I, J, K*, N3, P* and R1 haplogroups. Phylogeny of Y-chromosome haplogroups, and their frequencies in Hungarian and Szekler populations, are illustrated in Figure 4.

In both populations, the predominant Y-chromosome lineage was haplogroup R1. Haplogroup (Hg) R1 corresponds to the cluster defined by the M173 mutation. The M173-bearing Y-chromosomes in Europe are considered to delineate an ancient expansion from Central Asia during the Upper Paleolithic, ~30 000 years ago (Semino et al., 2000a; Underhill et al., 2001; Wells et al., 2001). About 50% of European Y-chromosomes share the M173 marker (Semino et al., 2000a). It has been found (Cruciani et al., 2002) that M173 chromosomes from Europe carry either M17/SRY_{10831} or M269 mutations, except in the case of three Danes reported by Sanchez et al. (2003).

Hungarians show a higher R-M17 haplogroup frequency than Szeklers (30% and 18.6% respectively), while only 15% of Hungarian Y chromosomes belong to the R-M269 lineage. In case of the Szeklers the frequency of the two clades is almost equal (18.6% R-M17 and 19.6% R-M269).

The presence of haplogroup P*(xM173) in Szeklers is striking, because this clade is otherwise almost absent in continental Europe (Semino et al., 2000a; Wells et al., 2001).

The second most prevalent haplogroup in both ethnic groups was Hg I, comprising 24.0% of the Hungarian and 21.7% of the Szekler Y-chromosomes, respectively. This clade is widespread all over Europe, but virtually absent elsewhere. The defining mutation of Hg I, M170, most likely arose in Europe before the Last Glacial Maximum (LGM) (Semino et al., 2000a). Among its major subclades, I1a-M253 shows the highest frequency in Northern Europe, whereas the I1b-P37 lineage peaks in Eastern Europe and in the Balkans (Rootsi et al., 2004). These two separate branches show divergent frequencies in the two ethnic groups; while I1a-M253 accounts for 16.5% of the Szekler Y-chromosomes, it is present in the Hungarian population with only 8% frequency. Haplogroup I1b-P37 occurs at 13% in Hungarians, but its frequency is only 5.2% in the Szekler population. Haplogroup I1b2 (M26) was not observed in our samples.

Interestingly, haplogroup J (12f2/M304) is widely distributed in both groups, with a frequency of 21.6% in Szekler and 16% in the Hungarian population. Haplogroup J-12f2 (M304), along with G-M201 and E-M35, first arrived in Europe from the Middle East in the Neolithic times (Semino et al., 2000a). The frequency of these lineages decreases from the Near East to Europe; they are more pronounced along the Mediterranean Coast than in continental Europe (Semino et al., 2000a, 2004).

Additional haplogroups with a frequency greater than 5% are: haplogroup E (10% in Hungarians; 9.2% in Szeklers, respectively), where E-M78 chromosomes account for almost all E representatives (89%; 90%), and haplogroup G-P15 in the Szekler population (5.2%).

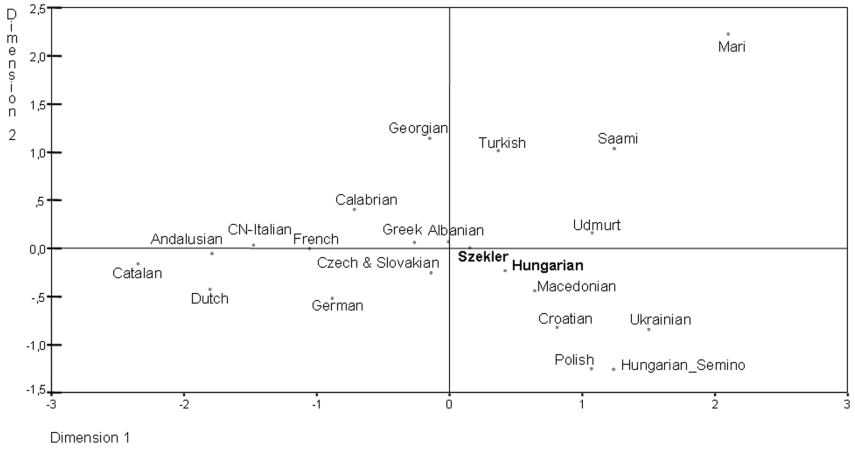


Figure 5 MDS analysis based on frequencies of the Y-chromosome haplogroups E-M35, G-M201, I-M170, J1-M267 (Eu 10), J2-M172, N3-Tat, R-M17 and R-M269 (Eu 18) in the examined Hungarian and Szekler groups (present paper) and in other European populations studied by Semino et al. (2000a).

Statistical Analysis of Y Chromosome Data

According to the intrapopulation diversity values (H), both of the populations are genetically heterogeneous (Hungarians H = 0.8089; Szeklers H = 0.8533).

By virtue of pairwise Fst genetic distance no significant difference was found between the Hungarian and Szekler populations (Fst < 0.05; Fst = 0.00348, P > 0.05; P > 0.23).

Pairwise values of Fst were also used to perform an MDS analysis (Figure 5), which shows that the Hungarian and Szekler populations are genetically closely related and are close to other populations from Central Europe and the Balkans.

Discussion

Analysis of Tat Polymorphism

The marker we examined, the T→C point mutation in the RBF5 locus, called the Tat C allele, is of interest in the Finno-Ugric context (Zerjal et al., 1997). It defines the Y chromosome haplogroup N3, which is present in northern and eastern Europe, but is virtually absent in the west and south. It is also frequent in northern Asia (Zerjal et al., 1997; Rootsi et al., 2000; Rosser et al., 2000; Semino et al., 2000a; Zerjal et al., 2001; Tambets et al., 2004), and is found at a high frequency in many Uralic-speaking populations (Lahermo et al., 1999; Rootsi et al., 2000; Tambets et al., 2001, 2004), including Voguls and Ostyaks who are the two closest linguistic cognates of Hungarians (Rédei, 2005); but it is hardly found in modern Hungarian populations (e.g. Tambets et al., 2004), even ethnic groups of Hungarians such as Csángós (Lahermo et al., 1999) or Palóc (Semino et al., 2000b).

These results immediately raise the question of whether the ancient Hungarians (early Magyars) who settled in the Carpathian Basin at the end of the 9th century (Kovács, 2005; Langó, 2005) and spoke a Uralic language, harboured this polymorphism or not.

Several explanations have been suggested for the striking absence of Tat C in the linguistically Uralic modern Hungarian population (Rootsi et al., 2000; Semino et al., 2000b). One is that Voguls and Ostyaks, and many other Siberian populations obtained this Y-chromosomal lineage only relatively recently, after the ancestors of the Magyars had left the Siberian forests for the great Eurasian steppe. Another is that Hungarians and Siberian Ugric-speaking populations have always been genetically unrelated despite their linguistic affinities. A third possibility is that the ancestral Magyars did have the Tat C allele, but lost it through genetic drift during their migration to Hungary, or after their settlement there.

To elucidate this question we attempted to screen 8 skeletal remains, from the age of the Hungarian Conquest, for the Tat C polymorphism, using bones from which ancient mtDNA fragments had been successfully recovered, which were therefore good candidates for Y-chromosomal analysis. Four of the 8 examined contained detectable Y-chromosomal DNA after whole genome amplification. Out of these, two possess the Tat C mutation. In case of the sample anc19 the presence of the ancestral Tat T allele was confirmed by typing the ancestral state (C allele) of the marker M9 (C→G) (YCC, 2002; Jobling & Tyler-Smith, 2003) (Figure 3). The fact that two of four ancient samples possessed the Tat C allele, is more than intriguing,

considering that from the 197 modern Hungarian-speaking males only one had this polymorphism (a finding consistent with previous studies (Lahermo et al., 1999; Rootsi et al., 2000; Rosser et al., 2000; Semino et al., 2000a, 2000b; Tambets et al., 2004).

The single observation of this particular lineage in the Szekler group may reflect some past contribution.

On the other hand, the low number of ancient samples has to be taken into consideration. The occurrence of this genetic variation in ancient Magyars could be the result of admixture or sharing a common male ancestor with other Uralic-speaking populations.

It seems highly probable that the nomadic Magyar tribes, when they entered Hungary, had preserved a very substantial Finno-Ugric male ancestry; substantially more distinctively Asian than the maternal mitochondrial ancestry we have described earlier (Tömöry et al., 2007). That would be consistent with the custom, observed in steppe societies from the earliest times, of slaughtering the males of other tribes given the opportunity, but carrying off their women.

Y Chromosome Variation in Modern Hungarian and Szekler Populations

In addition, 21 Y-chromosomal polymorphic markers of the two Hungarian-speaking populations were analysed (Figure 4). The Y-haplogroups observed in these two Hungarian ethnic groups are similar to those found among many other Europeans (Semino et al., 2000a).

Hg R1a1-M17 is the most frequent Y chromosome clade in the modern Hungarian population (30%) and encompasses 18.6% of the Szekler Y-chromosomal gene pool. The age of M17 has been estimated at 15 kilo years ago (KYA). It is likely that it arose in the Ukrainian glacial refuge, where this mutation is found with high frequency (>50%) (Semino et al., 2000a; Wells et al., 2001). This Hg is frequent in Eastern Europe, Central Asia and Northwest India (Semino et al., 2000a; Wells et al., 2001; Passarino et al., 2001; Kivisild et al., 2003).

The current distribution of R1a1-M17 in Europe shows an increasing west-east frequency and variance gradients with peaks among Finno-Ugric and Slavic speakers (Peričić et al., 2005 and references therein).

The R1a1-M17 frequency in Hungarians (30%) and Szeklers (18.6%) is comparable to that in their neighbours (e.g. Czechs and Slovaks, mainland Croatians, Bosnians, Romanians, Serbians) and some other Uralic-speaking populations (e.g. Estonian, Komis, Mordvin) (Tambets et al., 2004; Peričić et al., 2005 and references therein). The expansion of this haplogroup might have occurred at least three major episodes of gene flow: early post-LGM recolonization of Europe from the refugial area of present-day Ukraine (Semino et al., 2000a), migrations from the northern Pontic steppe between 3000 and 1000 BC and the historically attested Slavic migration from the 5th to 7th centuries AD (Peričić et al., 2005; Brather, 2001; Fusek, 2004).

Hg R1b-M269, in contrast to R1a1-M17, has its highest frequency in western Europe and decreases in eastern and southern Europe; the R1b variance shows multiple peaks in West Europe and Asia Minor (Semino et al., 2000a; Peričić et al., 2005). These spatial patterns possibly reflect the re-peopling of Europe from Iberia and Asia Minor during the Late Upper Paleolithic and Holocene (Cinnioğlu et al., 2004).

Similar frequencies of R1b as in the Hungarian speakers are found in some Slavic populations (mainland Croatians, Slovenians, Poles, Bulgarians); and in some Uralic-speakers (Komis, Khanties, Mordvin) as well as in Romanian and Turkish populations (Tambets et al., 2004; Peričić et al., 2005 and references therein).

The presence of central-Asian haplogroup P*(xM173) in Szeklers is unusual for a European population, since it is almost absent in continental Europe (Wells et al., 2001) and presumably reflects some Asian contribution, before or after reaching Transylvania.

Hg I-M170 is the only Y-chromosome haplogroup that is confined almost exclusively to the European continent (Semino et al., 2000a; Rootsi et al., 2004). Its virtual absence elsewhere, including the Near East, suggests that the defining mutation arose in Europe, most probably before the LGM (Semino et al., 2000a). Semino et al. (2000a) and Barać et al. (2003) proposed that the northern Balkans could have been a possible LGM refugium and a reservoir of M170. Rootsi et al. (2004) estimated that its major subclades, I1a, I1b and I1c all diverged from I* in the Late Upper Paleolithic/Mesolithic Period. Subclade I1a-M253 accounts for most of Hg I (88–100%) in Scandinavia, with a rapidly decreasing frequency toward both the East European Plain and the Atlantic fringe, but microsatellite diversity reveals that the Franco-Cantabrian refugial area could have been the source of the early spread of both I1a and the less common I1c (Rootsi et al., 2004).

Hg I1b-P37 is the most frequent haplogroup I clade in Eastern Europe and the Balkans (Rootsi et al., 2004). According to Peričić et al. (2005) the I1b*(xM26) lineages might have expanded from southeastern Europe to central, eastern and southern Europe, presumably not earlier than the Younger Dryas to Holocene transition and not later than the early Neolithic. In central and eastern Europe I1a and I1b show overlapping frequency gradients (Rootsi et al., 2004).

Haplogroup I was detected with almost equal frequency in the two modern populations: 24% in Hungarians and 21.7% in Szeklers. However, two of its major subclades-I1a-M253 and I1b* (xM26) – show an opposite occurrence in the two ethnic groups, 8% and 13%, respectively, in Hungarians, and 16.5% and 5.2% in Szeklers. These are within the range of normal central and eastern European values (Rootsi et al., 2004; Peričić et al., 2005). The elevated frequency of Hg I1a together with higher frequency of R1b-M269 in Szekler population might be the consequence, at least in part, of the genetic impact of people of German origin, who settled in Transylvania from the 12th century onwards (Transylvanian Saxons)(Makkai, 1990; Kristó, 2002).

Haplogroups J, E and G have been associated with the contribution of Neolithic farmers to the European gene pool (Semino et al., 2000a; Underhill et al., 2000). However, the phylogeographic analysis of haplogroups E and J revealed that multiple migratory events have spread these lineages during and after the Neolithic (Cruciani et al., 2004; Di Giacomo et al., 2004; Semino et al., 2004).

In the present study haplogroup J was unexpectedly common in the Hungarian-speaking populations (Hungarians: 16%, Szeklers: 21.6%). Haplogroup J, (defined by a 12f2 polymorphism –which is equivalent to M304-) (Cinnioğlu et al., 2004), is considered to have originated in the Middle East (Semino et al., 2004). The overall occurrence of the haplogroup displays an area of high frequencies (>20%) stretching from the Middle East to the central Mediterranean (Di Giacomo et al., 2004), whereas according to recent publications (Di Giacomo et al., 2004 and references therein) this haplogroup doesn't have a strong signature in the peoples of the northern Balkans and central Europe.

Hg J has two main clades; J1-M267 and J2-M172 (Cinnioğlu et al., 2004). J1-M267 has its highest frequency in the Middle East, North Africa and Ethiopia and its lowest in Europe (Semino et al., 2004). It has been suggested that Hg J1-M267 was spread by two temporally distinct migratory episodes, the most recent one probably associated with the diffusion of Arab people, mainly from the 7th century AD (Nebel et al., 2002; Semino et al., 2004). The J1-M267 Y-chromosomal lineage is notably frequent in Szeklers (10.3%; a value far above the range for other central and eastern European populations (Semino et al., 2000a, 2004; Di Giacomo et al., 2004), while its frequency in Hungarians (3.0%) is unremarkable.

Haplogroup J2-M172 is more prevalent in Europe than the J1 clade. Among its subclades J2e-M12 and J2f-M67 and their derivatives are found in Europe and in Asia (Semino et al., 2004). J2e-M12 is almost totally represented by its sublineage, J2e1-M102, which shows its maximum frequency in the southern Balkans and north-central Italy; it may have diffused from the former (Semino et al., 2004). J2f*-M67 is also a widespread clade which is most frequent in the Caucasus, whereas its derivative J2f1-M92 indicates affinity between Anatolia and southern Italy and is predominantly found in the northern Mediterranean from Turkey westward (Di Giacomo et al., 2004; Semino et al., 2004). J2f*-M67 and J2f1-M92 could have arrived in Europe from Anatolia across the Bosporus, as well as by seafaring Neolithic populations who may have reached southern Italy. The J2-M172* lineage displays a decreasing frequency gradient from the Near East toward western Europe. It has been proposed that the distribution of the J2-M172 (xM12) lineage is consistent with its spread to Europe through the Levantine corridor (Semino et al., 2004). Although J2-M172* encompasses most of the M172 Y-chromosomes in continental Europe and India, their degree of affinity and shared history remain uncertain. Semino et al. (2004) and Di Giacomo et al. (2004) point to Turkey and the Aegean as a relevant source for the J diversity observed throughout Europe.

Among these J2-M172 subclades, J2e1-M102 is more frequent in Szeklers (7.2%) than in Hungarians (4.0%), while the undifferentiated J2-M172* Y chromosomes are slightly more common in Hungarian population (8% vs. 3.1%). Both J2f*-M67 and J2f1-M92 lineages were detected in our study in one single individual, in each population.

The overall frequency of J2-M172 in the two populations (Hungarian: 13%; Szekler: 11.3%) is in the same order of magnitude as reported for Czechs and Slovaks (Semino et al., 2000a), Romanians, Bulgarians (Di Giacomo et al. 2004) and Ukrainians (Semino et al. 2004).

The elevated overall frequency of haplogroup J in the Szekler samples (21.6%) might be partially attributed to genetic drift, since Szeklers lived in relative isolation as a socially exclusive population in the mountains of Transylvania from the Middle Ages (12th century) until the 19th century (Makkai 1990; Kristó 2002). But that cannot explain the even higher frequency in Hungarians. On the other hand, the elevated frequency of J in both groups could also be due to a range of historical events.

One is the expansion of the Ottoman Empire from the 16th century AD; refugees from the Balkan area fled to Hungarian territory (Pálffy 2000).

Another is in accordance with our mtDNA analysis of ancient Hungarians (Tömöry et al. 2007); populations who lived in close contact with the early Magyars during their migration from the Ural region to the Carpathian Basin left substantial imprints in their gene pool. In this context, one should note that statistical analysis of mtDNA variation of ancient Hungarian and modern Eurasian populations

shows that the early Magyars were genetically close not only to some populations from Central Asia and Europe, but also to some of the Near-East populations – Syrian, Palestinian, Turkish – where Y chromosome haplogroup J is present with high frequency (Semino et al. 2000a, 2004; Cinnioğlu et al. 2004; Di Giacomo et al. 2004; Flores et al. 2005).

However, the possibility that the J cluster was present in the ancient Hungarian population and may have left detectable traces in the paternal gene pool of modern Hungarian and Szekler males deserves future study. At this stage of analysis other explanations, such as genetic drift or more recent gene flow are plausible.

There is a discrepancy between the frequencies of haplogroup J reported here and the results of some previous studies (Rosser et al. 2000; Semino et al. 2000a, 2000b, 2004), in which it was found to be rare (2–3%) in modern Hungarian-speaking populations; (but conversely, in a Hungarian sample studied by Rootsi et al. (2000) it was detected at a level of 14%). The possible reason for this discrepancy might lie in differential sampling. Hungarian samples included in this paper were collected from different parts of Hungary, but the most represented area was the Great Hungarian Plain (90 out of 100 individuals were born there). Hungarian samples studied by Semino et al. (2000a), however, originated from subjects from Budapest and other northern regions in Hungary. The most frequently represented were Palócs (Semino personal communication), an ethnic minority in Hungary, from a border region of the northern Carpathian Basin, neighbours of the Slavs in the higher mountains, with a distinctive Hungarian dialect (Kósa 1998; Semino et al. 2000b).

The MDS plot (Figure 5) based on Fst values of population data from Semino et al. (2000a) and those of this study, shows that the Hungarian and Szekler populations are genetically closely related, and close also to other populations from Central Europe and the Balkans. However, the second dimension clearly separates our samples from those Hungarians studied by Semino et al. (2000a). The latter group has closer genetic relationships with Polish and Ukrainian populations. A possible explanation might be that in the 19th century extensive admixture occurred between inhabitants of northern parts of Hungary and those of neighbouring northern regions (Kósa 1998). This might also be responsible for the large proportion, 60%, of R1a1-M17 Y chromosomes reported (Semino et al., 2000a).

Haplogroup E3b-M35 occurs at 10% frequency in Hungarians and 9.2% in Szeklers with E3b1-M78 chromosomes accounting for almost all representatives (~90%). Hg E is mainly African, but its clade E3b-M35 has also been observed in Europe, where it is believed to have arrived during and after Neolithic. Both E3b-M35 and its derivative (E3b1-M78) probably originated in eastern Africa (Semino et al. 2004; Cruciani et al. 2004). Hg E3b1-M78 has been observed over a wide area, including eastern and northern Africa, the Near East and Europe, where it represents by far the most common E3b subhaplogroup (Cruciani et al. 2004; Semino et al. 2004). This lineage shows its highest frequency in southern and southeastern Europe and declining frequencies toward western, central and eastern Europe (Peričić et al. 2005). The frequencies of E3b1-M78 in our samples (Hungarian: 9%; Szekler: 8.2%) are within the range of central and eastern European values (Peričić et al. 2005). According to Cruciani et al. (2004) the main contributor to the present distribution of E3b lineages in Europe was an expansion from the Balkans to western and southern-central Europe (~7.8 KYA). Peričić et al. (2005) suggested that one of the major routes for E3b1-M78 expansion was the Vardar-Morava-Danube river system leading from south and southeastern to continental Europe.

Conclusion

Our data suggest that the Tat C allele, which is widespread in Uralic-speaking populations, was substantially present in the ancient Magyar population when they crossed the Carpathians and settled in the Carpathian Basin. Our findings provide further evidence for its virtual absence in recent Hungarian-speaking populations, with the exception of a single male in the Szekler group. This contrast, despite the relative linguistic stability, may be attributed to a combination of the Magyars being a dominant elite, whose language was accepted by the more numerous pre-existing populations (mostly Slavs and Avars), and of the effects of a number of substantial post-Magyar immigrations and incursions.

The Y-chromosomal patterns of the modern Hungarians and Szeklers can for the most part be adequately explained within the European paternal genetic landscape. As with other Europeans, the Y chromosomes are characterized by early lineages derived from Paleolithic inhabitants, and by a minor impact of Neolithic and post-Neolithic migratory episodes. Consistent with previous studies, Hungarian-speaking populations are genetically closely related to their geographic neighbours. The Hungarian and Szekler groups cluster together with some other central Europeans (e.g. Czechs and Slovaks), but mainly with Balkan populations.

There are two exceptions. Haplogroup P*(xM173) is almost absent in continental Europe. The presence of this haplogroup in the Szeklers may indicate a connection with Central Asian populations. Also, there is an elevated haplogroup J frequency. This may reflect Anatolian and southern Balkan contributions to the gene pools of Hungarians and Szeklers, but historical data and the comparative analyses of maternal lineages of ancient Hungarian

population suggest that the earlier migrations of the Magyars may also have contributed to the presence of this lineage in the Carpathian Basin.

Acknowledgements

We thank Mrs Mária Radó and Mrs Gabriella Lehõcz for skilled technical assistance; Imre Cserpán, István Németh and Zsolt Pénzes for their professional advice.

This work was supported by grants 5/088/2001 and OM-00050/2004 of the Hungarian National Research and Development Programs.

Electronic Information:

The URL for dCAPS Finder 2.0 program is as follows: http://helix.wustl.edu/dcaps/dcaps.html.

References

Barać, L., Peričić, M., Klarić, I. M., Rootsi, S., Janićijević, B., Kivisild, T., Parik, J., Rudan, I., Villems, R. & Rudan, P. (2003) Y chromosomal heritage of Croatian population and its island isolates. *Eur J Hum Genet* **11**, 535–42.

Bálint, Cs. (1996) A kora középkori kelet-európai steppe régészete és a 9–10. századi magyarok. (The Archaeology of the Early central European Steppe and the Hungarians of the 9th and 10th Centuries.) *MT* **103**, 937–947.; 1041. Akadémiai Kiadó.

Bálint, Cs. (2005) Ki volt "magyar" a honfoglalás korban és Szent István korában? In: Romsics, I., Szegedy-Maszák, M. (eds.) *Mi a magyar?* pp. 37–56. Rubicon Kiadó: Budapest.

Bálint, Cs. (2006) Az etnosz a korai középkorban. (A kutatás lehetségei és gátjai.) (The *Ethnos* in the Early Middle Ages. Possibilities and Limits of Research.) *Sz* **140**, 277–347. Magyar Történeti Társulat.

Beer, Zs., Csete, K. & Varga, T. (2004) Y-chromosome STR haplotype in Szekely population. *Forensic Sci Int* **139**, 155–58.

Belgrader, P., Marino, M. M., Lubin, M. & Barany, F. (1996) A multiplex PCR-ligase detection reaction assay for human identity testing. *Genome Sci Technol* **1**, 77–87.

Bogácsi-Szabó, E., Kalmár, T., Csányi, B., Tömöry, Gy., Czibula, Á., Priskin, K., Horváth, F., Downes, C. S. & Raskó, I. (2005) Mitochondrial DNA of ancient Cumanians: culturally Asian steppe nomadic immigrants with substantially more western Eurasian mitochondrial DNA lineages. *Hum Biol* **77**, 639–62.

Bóna, I. (1990) Die ungarische Landnahme. Siebenbürger und der Ostrand der Großen Ungarischen Tiefebene zwischen 895 und 1003. In: Köpeczi, B. (Ed.). *Kurze Geschichte Siebenbürgens*. pp. 109–136.; 697–700. Akademia Verlag: Budapest.

Brather, S. (2001) *Archäologie der westlichen Slawen*. Reallexikon der Germanischen Altertumskunde – Ergänzungsbände, 30. Walter de Gruyter Verlag: Berlin–New York.

Cinnioğlu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L., Lillie, A. S., Roseman, C. C., Lin, A. A., Prince, K., Oefner, P. J., Shen, P., Semino, O., Cavalli-Sforza, L. L. & Underhill, P. A. (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* **114,** 127–48.

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D. C., Oefner, P. J., Torroni, A., Cavalli-Sforza, L. L., Scozzari, R. & Underhill, P. A. (2002) A back migration from Asia to sub Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* **70**, 1197–1214.

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E. B., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Calì, F., Akar, N., Richards, M., Torroni, A., Novelletto, A. & Scozzari, R. (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* **74**, 1014–22.

Czeizel, A. E., Benkman, H. G. & Goedde, H. W. (1991) *Genetics of the Hungarian populations*. Springer-Verlag: Berlin.

Di Giacomo, F., Luca, F., Popa, L. O., Akar, N., Anagnou, N., Banyko, J., Brdicka, R., Barbujani, G., Papola, F., Ciavarella, G., Cucci, F., Di Stasi, L., Gavrila, L., Kerimova, M. G., Kovatchev, D., Kozlov, A. I., Loutradis, A., Mandarino, V., Mammi', C., Michalodimitrakis, E. N., Paoli, G., Pappa, K. I., Pedicini, G., Terrenato, L., Tofanelli, S., Malaspina, P. & Novelletto, A. (2004) Y chromosomal haplogroup J as a signature of the post-Neolithic colonization of Europe. *Hum Genet* **115**, 357–71.

Flores, C., Maca-Meyer, N., Larruga, J. M., Cabrera, V. M., Karadsheh, N. & Gonzalez, A. M. (2005) Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. *J Hum Genet* **50**, 435–41.

Fodor, I. (1996) The Hungarian Conquest. In: Fodor, I., Révész, L., Wolf, M., Nepper, I. M. (Ed.). *The Ancient Hungarians*. pp. 13–18. Hungarian National Museum: Budapest.

Fusek, G. (2004) "Slawen" oder Slawen? Eine polemische Auseinandersetzung über eine wertvolle Monographie. *SlovArch* **52**, 161–181. Arheologicky Ústav SAV.

Guglielmino, C. R., Piazza, A., Menozzi, P. & Cavalli-Sforza, L. L. (1990) Uralic genes in Europe. *Am J Phys Anthropol* **83**, 57–68.

Guglielmino, C. R., De Silvestri, A. & Béres, J. (2000) Probable ancestors of Hungarian ethnic groups: an admixture analysis. *Ann Hum Genet* **64**, 145–159.

Hammer, M. F., Redd, A. J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M. A., Jenkins, T., Ostrer, H. & Bonné-Tamir, B. (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc Natl Acad Sci U S A* **97**, 6769–74.

Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., Utermann, G., Sykes, B. & Pääbo, S. (1994) Molecular genetic analyses of the Tyrolean Ice Man. *Science* **264**, 1775–78.

Hanson, E. K. & Ballantyne, J. (2005) Whole genome amplification strategy for forensic genetic analysis using single or few cell equivalents of genomic DNA. *Anal Biochem* **346**, 246–57.

Jobling, M. A. & Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4**, 598–612.

Kalmár, T., Bachrati, Cs.Z., Marcsik, A. & Raskó, I. (2000) A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Res* **28**, E67.

Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R. S., Redd, A. J., Zegura, S. L. & Hammer, M. F. (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* **69**, 615–28.

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H. V., Stepanov, V., Gölge, M., Usanga, E., Papiha, S. S., Cinnioğlu, C., King, R., Cavalli-Sforza, L. L., Underhill, P. A. & Villems, R. (2003) The genetic

heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* **72**, 313–32.

Kordé, Z. (1994) Székelyek. In: Kristó, Gy. (Ed.), *Korai magyar történeti lexikon (9–14. század)*. pp. 623–25. Akadémiai Kiadó: Budapest.

Kósa, L. (1998) '*Ki népei vagytok?*'. Magyar néprajz: Budapest.

Kovács, L. (2005) Remarks on the archaeological remains of the 9th–10th century Hungarians. In: Mende B. G. (Ed.) *Research on the prehistory of the Hungarians: A review VAH* **18**: 351–68. Hungarian Academy of Sciences Archaeological Institute.

Kristó, Gy. (2002) *A székelyek eredete*. Balassi Kiadó: Budapest.

Kristó, Gy. (2005) *A korai Erdély*. Szegedi Tudományegyetem Középkorász Mhelye: Szeged.

Lahermo, P., Savontaus, M. L., Sistonen, P., Béres, J., de Knijff, P., Aula, P. & Sajantila, A. (1999) Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur J Hum Genet* **7**, 447–58.

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

Langó, P. (2005) Archaeological research on the conquering Hungarians: A review. In: Mende, B. G. (Ed.) *Research on the prehistory of the Hungarians: A review VAH* **18**: 175–340. Hungarian Academy of Sciences Archaeological Institute.

Makkai, L. (1990) Siebenbürgen im mittalterlichen Königreich Ungarn. In: Köpeczi, B. (Ed.) *Kurze Geschichte Siebenbürgens*. pp. 137–240. Akademia Verlag: Budapest.

Mesterházy, K. (1997) A honfoglaló magyarok tárgyi emlékei. *Életünk* pp. 30–67.

Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215.

Nebel, A., Landau-Tasseron, E., Filon, D., Oppenheim, A. & Faerman, M. (2002) Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. *Am J Hum Genet* **70**, 1594–96.

Neff, M. M., Turk, E. & Kalishman, M. (2002) Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet* **18**, 613–15.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press: New York, NY, USA.

Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L. & Tagliabracci, A. (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci Int* **157**, 23–35.

Pálffy, G. (2000) *A tizenhatodik század története*. Pannonica Kiadó: Budapest.

Paracchini, S., Arredi, B., Chalk, R. & Tyler-Smith, C. (2002) Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res* **30**, E27.

Passarino, G., Semino, O., Magri, C., Al-Zahery, N., Benuzzi, G., Quintana-Murci, L., Andellnovic, S., Bullc-Jakus, F., Liu, A., Arslan, A. & Santachiara-Benerecetti, A. S. (2001) The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. *Hum Immunol* **62**, 922–32.

Peričić, M., Lauc, L. B., Klarić, I. M., Rootsi, S., Janićijević, B., Rudan, I., Terzić, R., Čolak, I., Kvesić, A., Popović, D., Šijački, A., Behluli, I., orević, D., Efremovska, L., Bajec, . D., Stefanović, B. D., Villems, R. & Rudan, P. (2005) High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Mol Biol Evol* **22**, 1964–75.

Rédei, K. (2005) *történetünk kérdései. A nyelvészeti dilettantizmus kritikája*. Balassi Kiadó: Budapest.

Révész, L. (1996) *A karosi honfoglalás kori temetk. Adatok a Fels-Tisza-vidék X. századi történetéhez*. (Die Gräberfelder von Karos aus der Landnahmezeit. Archäologische Angaben zur Geschichte des Oberen Theiβgebietes im 10. Jahrhundert.) Magyarország honfoglalás kori és kora Árpád-kori sírleletei, 1. Miskolc. Herman Ottó Museum.

Reynolds, J., Weir, B. S. & Cockerham, C. C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**, 767–79.

Róna-Tas, A. (1995) *A magyarság korai története*. Magyar történeti Könyvtár, 9. University of Szeged: Szeged.

Róna-Tas, A. (1999) *Hungarians and Europe in the Early Middle Ages*. Central Europe University Press: Budapest.

Rootsi, S., Kivisild, T., Tambets, K., Adojaan, M., Parik, J., Reidla, M., Metspalu, E., Laos, S., Tolk, H. V. & Villems, R. (2000) On the phylogeographic context of sex-specific genetic markers of Finno-Ugric populations. In: Kiinnap, A. (Ed.) *The roots of peoples and languages of Northern Eurasia II and III*. pp. 148–164. University of Tartu, Division of Uralic Languages/Societas Historiae Fenno-Ugricae:Tartu.

Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barać, L., Peričić, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., 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, P. J., Shen, P., Beckman, L., Mikerezi, I., Terzić, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Benerecetti, A. S., 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**, 128–37.

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Côrte-Real, H. B. S. M., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kučinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A. V., Mitchell, R. J., Nafa, K., Nicholson, J., Nørby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previderé, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C. & Jobling, M. A. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**, 1526–43.

Sanchez, J. J., Børsting, C., Hallenberg, C., Buchard, A., Hernandez, A. & Morling, N. (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Sci Int* **137**, 74–84.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin ver. 2.000: A software for population genetics data analysis*.Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D.,

Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L. & Underhill, P. A. (2000a) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–59.

Semino, O., Passarino, G., Quintana-Murci, L., Liu, A., Beres, J., Czeizel, A. & Santachiara-Benerecetti, A. S. (2000b) MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. *Eur J Hum Genet* **8**, 339–46.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L. L., Underhill, P.A. & Santachiara-Benerecetti, A. S. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74**, 1023–34.

Sindbæk, S. M. (1999) A Magyar Occurrence. Process, practice and ethnicity between Europe and the Steppes. *ActaArch* **70**, 149–64.

Szíj, E. (2005) Research on the prehistory of the Hungarians and Finno-Ugric studies. In: Mende, B. G. (Ed.) *Research on the prehistory of the Hungarians: A review*. *VAH* **18**: 115–56. Hungarian Academy of Sciences Archaeological Institute.

Szõke, B. (1962) *A honfoglaló és kora Árpád-kori magyarság régészeti emlékei*. *RégTan 1*. Akadémiai Kiadó: Budapest.

Tambets, K., Rootsi, S., Kivisild, T. & Villems, R. (2001) The concepts of Richard Indreko about the origin of the Finno-Ugric speakers and the population genetics of the extant North-East European populations. *TRAMES* **5** (55/50), 1, 59–74.

Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E. L., Tolk, H. V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J. P., Barać, L., Peričić, M., Rudan, P., Terzić, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G. F., Anagnou, N., Pappa, K. I., Michalodimitrakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L. & Villems, R. (2004) The western and eastern roots of the Saami–the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* **74**, 661–82.

Tömöry, Gy., Csányi, B., Bogácsi-Szabó, E., Kalmár, T., Czibula, Á., Csõsz, A., Priskin, K., Mende, B., Langó, P., Downes, C. S. & Raskó, I. (2007) Comparison of maternal lineage and biogeographic analysis of ancient and modern Hungarian populations. *Am J Phys Anthropol* **134**, 354–68.

Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* **7**, 996–1005.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–61.

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon, Lahr, M., Foley, R. A., Oefner, P. J. & Cavalli-Sforza, L. L. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* **65**, 43–62.

Walsh, P. S., Metzger, D. A. & Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506–13.

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A. & Bodmer, W. F. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* **98**, 10244–49.

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339–48.

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F. R., Schiefenhovel, W., Fretwell, N., Jobling, M. A., Harihara, S., Shimizu, K., Semjidmaa, D., Sajantila, A., Salo, P., Crawford, M. H., Ginter, E. K., Evgrafov, O. V. & Tyler-Smith, C. (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* **60**, 1174–83.

Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A. V., Krumina, A., Kučinskas, V., Hurles, M. E. & Tyler-Smith, C. (2001) Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol Biol Evol* **18**, 1077–87.

Received: 4 June 2007
Accepted: 2 February 2008