Heredity (2008) 101, 416–419

www.nature.com/hdy

ORIGINAL ARTICLE

H1 tau haplotype-related genomic variation at *17q21.3* as an Asian heritage of the European Gypsy population

PZ Álmos[1,2], S Horváth[1], Á Czibula[2], I Raskó[2], B Sipos[2], P Bihari[2], J Béres[3], A Juhász[1], Z Janka[1] and J Kálmán[1]

[1]*Department of Psychiatry, Faculty of Medicine, University of Szeged, Szeged, Hungary;* [2]*Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary and* [3]*Department of Hungarian Congenital Abnormality and Rare Disease Registry, National Centre For Healthcare Audit and Improvement, Budapest, Hungary*

In this study, we examine the frequency of a 900 kb inversion at *17q21.3* in the Gypsy and Caucasian populations of Hungary, which may reflect the Asian origin of Gypsy populations. Of the two haplotypes (H1 and H2), H2 is thought to be exclusively of Caucasian origin, and its occurrence in other racial groups is likely to reflect admixture. In our sample, the H1 haplotype was significantly more frequent in the Gypsy population (89.8 vs 75.5%, $P<0.001$) and was in Hardy–Weinberg disequilibrium ($P=0.017$). The *17q21.3* region includes the gene of microtubule-associated protein tau, and this result might imply higher sensitivity to H1 haplotype-related multifactorial tauopathies among Gypsies. *Heredity* (2008) **101,** 416–419; doi:10.1038/hdy.2008.70; published online 23 July 2008

Keywords: Gypsy; Roma; genomic variation; tau haplotype; tauopathies

Introduction

Recent studies have estimated that at least 10% of the human genome is polymorphic for structural rearrangements. New developments in genome-scanning array and comparative DNA sequencing technologies have made possible the characterization and assessment of all classes of genomic variation, including structural variation. These developments provide an opportunity to generate geographical maps of the frequency of structural variants. Some of them are typical for different ethnic groups or for certain populations (Gu *et al*., 2007; O'Hara, 2007; Spielman *et al*., 2007). The presence of these variations is thought to be an important contributor to the evolution in human genetic diversity and can generate difference in disease susceptibility (Feuk *et al*., 2006).

One of the most notable structural variants found to date is a 900 kb inversion, which is located to *17q21.3* (Stefansson *et al*., 2005). This region includes the gene of microtubule-associated protein tau (MAPT), which is widely studied, as it contributes to several human diseases (Hardy *et al*., 2006). These extensive investigations have revealed that two main nonrecombining MAPT locus haplotypes (H1 and H2) can be distinguished. The H1 haplotype and/or subhaplotypes play a general role in the development of sporadic tauopathies (Laws *et al*., 2007; Myers *et al*., 2007). It has also become obvious that there are notable differences in the geographical distribution of the MAPT-related haplotypes. The H2 haplotype is rare in Africans, and almost absent in East Asians and Native Americans, but very frequent (20–30%) in populations of Caucasian origin (Evans *et al*., 2004; Stefansson *et al*., 2005). It has even been postulated that the H2 haplotype was contributed to the human genome by *Homo neanderthalensis* (Hardy *et al*., 2005).

In this study, we examined the distribution in two historically and ethnically different populations (Gypsies and Caucasian Hungarians) from the same geographical area. The Caucasian Hungarians belong to the Uralic linguistic family, a diverse group of people related by an ancient common linguistic heritage, distinct from that of the Indo-European speakers who surround them. Of the approximately 25 million Finno-Ugrians, the best known are the Estonians, the Finns and the Hungarians. Around the 5th century BC, the ancient Hungarians were caught up in a wave of migrations that swept the steppes and were displaced from their western Siberian homeland. Migrating westwards, the Hungarians arrived in 895 in the Carpathian Basin, an area where the overwhelming majority of the indigenous population was Slavic (Figure 1). Various genetic appraisals have estimated that the newly arrived Hungarians accounted for 10–50% of the total population of the Carpathian Basin (Cavalli-Sforza, 1994). During the turbulent history of present-day Hungary, the mixing process has continued, and Hungarians can now be regarded as members of a mixed European population (Semino *et al*., 2000).

In contrast to Hungarians, Gypsies are a conglomerate founder population with Asian roots, imbedded in a genetically different Caucasian population. The social

Correspondence: Dr PZ Álmos, Department of Psychiatry, Faculty of Medicine, University of Szeged, Semmelweis u. 6. Szeged H-6720, Hungary.
E-mail: almos@nepsy.szote.u-szeged.hu
Received 16 May 2007; revised 17 May 2008; accepted 20 June 2008; published online 23 July 2008

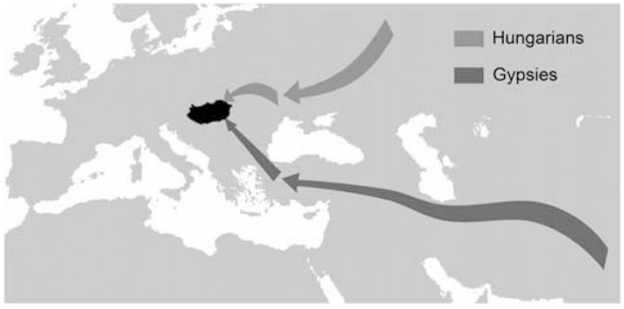

Figure 1 Migration routes of Gypsies from Punjab region and Hungarians from Uralic Mountains.

sciences and comparative linguistic studies have hinted at the Asian origin, and this has been supported by population genetic studies of single-locus polymorphism, of multi-locus STR Y chromosome haplotypes and of mtDNA haplotypes (Gresham *et al*., 2001; Kalaydjieva *et al*., 2001a; Morar *et al*., 2004).

The combined evidence suggests that Gypsies migrated from the Punjab region of northwest India 1000–1500 years ago and traveled through Asia (along Persia, today's Armenia and Turkey). The main stream moved into the Balkans and Greece and some of them into eastern Europe ahead of the Turks (Figure 1). Early diaspora appeared in western Europe around the period from the fourteenth to the fifteenth century, and another wave of migrations to western Europe started after the abolition of serfdom in the Habsburg Empire in 1841, and recently from 1989 after the disappearance of the Iron Curtain (Kalaydjieva *et al*., 2005).

At present, 8–10 million Gypsies live in fragmented subisolates in Europe, approximately 600 000 of them in Hungary. In Gypsy society, the primary unit is the group, and groups are members of metagroups. They live in a closed society structure, with rare admixture with other populations, and a relatively high rate of consanguinity (Assal *et al*., 1991). There appears to have been population bottlenecks, both when they left India and during the European segregation. A high intragroup diversity can be observed (Gresham *et al*., 2001).

Hungarian Gypsies were not classified in previous publications or were included among western European Gypsies (Morar *et al*., 2004). However, we think that the comparison of the Hungarian Gypsy population is an adequate choice for genetic investigations because the ethnic diversity in Hungary is not as high as in the Balkans, and it is possible to distinguish three well-described metagroups among Hungarian Gypsies. Carpathian Gypsies or Romungros are the least characterized and intact metagroup. Their language consists elements from Beas, Lovari and Hungarian. They represent the 70% of the Gypsies living in Hungary. The two smaller metagroups are more closed and cohesive; they live typically in separated parts of smaller villages or towns. They preserve their traditions and language; as a consequence, the assimilation with other metagroups or with the Caucasoid Hungarian population is low. Beas represents 10% of the Hungarian Gypsy population; their migration to the Carpatian Basin came from the Central-West Balkans. They speak the Beas language. The Olahs, with a proportion of 20% from the Hungarian Gypsy population, arrived at the Carpatian

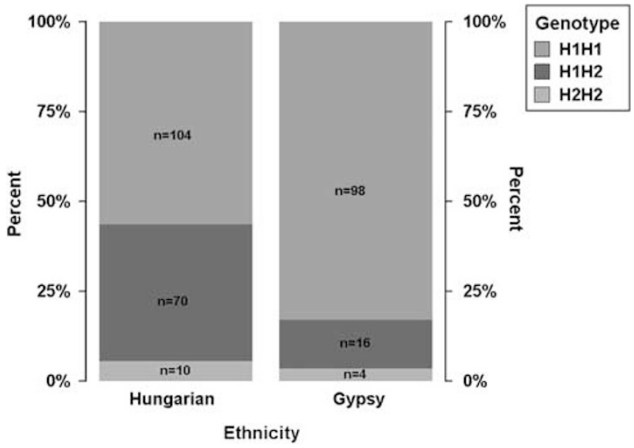

Figure 2 The distribution pattern of MAPT genotypes.

Basin from the territory of today's Romania and they speak the Lovari language. They are the descendants of the Walachian/Vlax Gypsies, the most studied Gypsy population (Kalaydjieva *et al*., 2001a).

In this study, our goal was to evaluate the H1–H2 haplotype frequencies in the populations mentioned above by using a polymorphism of MAPT gene as a marker.

Materials and methods

Sample characteristics

In this study, 118 healthy Gypsies of the Olah/Vlax metagroup and 184 healthy Caucasian Hungarians were genotyped (Figure 2). The Gypsy participants were recruited from three villages in the same geographical area in northeastern Hungary. The Hungarians were employees and students of Department of Psychiatry, University of Szeged, and Department of Hungarian Congenital Abnormality and Rare Disease Registry of the National Centre For Healthcare Audit and Improvement and their acquaintances, who were matched with the Gypsy volunteers for age and gender. After complete description of the study to the subjects, written informed consent was obtained.

DNA isolation

The genomic DNA of Gypsy and control subjects was extracted from peripheral blood according to a standard method (Davies, 1986).

PCR amplification

The selected region was amplified means of the PCR. The inverted chromosome region was screened by using the standardly used biallelic intron 9 deletion-inversion polymorphism (Baker *et al*., 1999). The following primer pairs were used: forward: 5′-GAAGACGTTCTCACTG ATCTG-3′; reverse: 5′-AGGAGTCTGGCTTCAGTCTC-3′.

Polymerase chain reaction amplification was carried out in 20 μl reaction volume containing 2 μl of 10 × ZenonBio, 10 × reaction buffer, 50 nM of each of the primers, 0.5 μM of each of the dNTPs, 4 mM $MgCl_2$, 100 ng of DNA extract and 0.3 U of ZenonBio TaqPolymerase. The amplification protocol was as follows: 3 min at 93 °C, 30 cycles of 93 °C for 60 s, 60 °C for 60 s and 72 °C

for 60 s, and final extension at 75 °C for 5 min. A volume of 7 µl of PCR product was run on 6% native polyacrylamide gel and visualized after ethidium bromide staining by UV transillumination, and the size of the products was determined with the GelBase gel documentation system (UVP).

Statistical analysis

The departure from the Hardy–Weinberg equilibrium was tested by using the 'HWE.test' function (*P*-value calculated by the exact method) of the genetics R package (R version 2.4.0, R Development Core Team, 2006; Warnes, 2008). Fisher's exact tests carried out in R were used to determine the significance of differences in genotype and allele frequencies.

Results

The MAPT allele frequencies in the Caucasian sample were in Hardy–Weinberg equilibrium ($P = 0.842$). A deviation from the Hardy–Weinberg equilibrium was observed in the Gypsy population sample ($P = 0.017$).

The distribution of MAPT genotypes are presented in Figure 2. The MAPT H1 homozygote haplotype is seen to be overrepresented in the Gypsies as compared with the Caucasians (83.0% ($n = 98$) vs 56.5% ($n = 104$) (one-tailed $P < 0.001$). H1/H2 heterozygotes prevail in the Caucasian population (38.0%; $n = 70$ in the Caucasians vs 13.6%; $n = 16$ in the Gypsies) (one-tailed $P < 0.001$). The calculated frequency of the H1 allele in the Gypsy population was greater than that in the Caucasians (89.8% ($n = 212$) vs 75.5% ($n = 278$) (one-tailed $P < 0.001$), whereas H2 allele was more dominant in the Caucasian population.

Discussion

Our results indicate a different proportion of the inversion at *17q21.3* in Hungarian Olah Gypsies as compared with Caucasian Hungarians. This study has revealed that Hungarian Olah Gypsies, who are related to the Asian population, carry the H1 allele at a higher proportion than Caucasian populations. This supports the notion that *17q21.3* structural variation and tau haplotypes are suitable markers for the demonstration of the degree of admixture in a well-characterized non-Caucasian population.

The 23.8% H2 allele frequency in the Hungarian population accords well with the frequency of ~25% in middle-eastern and European populations (Evans *et al*., 2004). The previously reported 8% of H2 allele frequency (Evans *et al*., 2004) in the Finnish population stands closer to the Asian genotype distribution. These results suggest that the Finnish population experienced less admixture than the Caucasian population of Hungary, and the Asian descent of the latter is not detectable by this method.

In our Gypsy sample, the frequency of the H1 allele was lower than previous estimates from populations of Asian origin (only populations from South Pakistan were similar) (Evans *et al*., 2004). The lower frequency of the H1 haplotype in the Gypsy population may be a consequence of their coexistence for centuries and partial admixture with H2 carrier Caucasian populations. This effect is likely to have been strengthened by the fact that the Olah/Vlax metagroup traditionally tolerates marriages with non-Gypsy women, whereas some other Gypsy groups do not. The deviation from the Hardy–Weinberg equilibrium in the Gypsy group can be explained by the population genetic effect of their closed society structure and the higher rate of consanguineous mating.

The Gypsy ethnic group was ignored for centuries by Western society and medicine. The United Nations Development Programme (www.undp.org) and the Decade of Roma Inclusion 2005–2015 (www.romadecade.org) recognized the importance of medical and social studies. In the past decade, various Mendelian diseases with a carrier rate of 5–15% have been identified in the Gypsy population (Kalaydjieva *et al*., 2001b), but multifactorial tauopathies have not been well described in Gypsies. This can be explained by their social and medical neglect and the fact that tauopathies are typically late-onset neurodegenerative diseases, although the average life expectancy of Gypsies is 10–15 years lower than the European standard (Sepkowitz, 2006).

H1 carriers are under a negative selection, as H2 carrier women have more children (Stefansson *et al*., 2005; Voight *et al*., 2006) and because of the possible role of H1 allele in tauopathies. Alzheimer's disease (Myers *et al*., 2005; Laws *et al*., 2007), Parkinson's disease (Skipper *et al*., 2004), progressive supranuclear palsy (Pittman *et al*., 2004), argyrophilic grain disease (Fujino *et al*., 2005), corticobasal degeneration (Buee and Delacourte, 1999) and the Parkinson–dementia complex of Guam (Sundar *et al*., 2007) are all associated with MAPT H1. In addition, it seems that, besides carrying H1 allele, there are other factors that influence disease onset. Differences in gene expression or in alternative splicing or both may lead to enhanced tangle formation and the development of the disease (Avila, 2006; Caffrey *et al*., 2006; Hardy *et al*., 2006; O'Hara, 2007). Exposure to different (European) environmental factors means differences in epigenetic effects on gene expression (Spielman *et al*., 2007). For instance, a recent study (Winkler *et al*., 2007) indicated H1/H1 genotype as an ethnically dependent risk factor of Parkinson's disease, and another one raised remarkable suggestions on this field (Fung *et al*., 2005). An early work also observed association regarding tau variants and Asian versus Caucasian populations in progressive supranuclear palsy (Conrad *et al*., 1998). Thus, higher H1 frequency in Gypsies might be a risk factor of multifactorial disorders and be manifested as an elevated susceptibility to tauopathies among the Gypsy population in Europe. Further investigations are needed in populations with high H1 frequency where the social and medical aspects and the average life expectancy are better.

Acknowledgements

We thank all probands for their participation in this study. We are grateful for the help of Zsolt Pénzes in the statistical analysis. This work was supported by a grant from OTKA 5K526/2005 (Hungarian Scientific Research Fund). SH was a Bolyai Scholar during part of this project.

References

Assal S, Susanszky E, Czeizel A (1991). High consanguinity rate in Hungarian gypsy communities. *Acta Paediatr Hung* **31**: 299–304.

Avila J (2006). Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett* **580**: 2922–2927.

Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J *et al.* (1999). Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet* **8**: 711–715.

Buee L, Delacourte A (1999). Comparative biochemistry of tau in progressive supranuclear palsy, corticobasal degeneration, FTDP-17 and Pick's disease. *Brain Pathol* **9**: 681–693.

Caffrey TM, Joachim C, Paracchini S, Esiri MM, Wade-Martins R (2006). Haplotype-specific expression of exon 10 at the human MAPT locus. *Hum Mol Genet* **15**: 3529–3537.

Cavalli-Sforza LL, Piazza A (1994). *The History and Geography of Human Genes*. Princeton University Press: Princeton, New Jersey.

Conrad C, Amano N, Andreadis A, Xia Y, Namekataf K, Oyama F *et al.* (1998). Differences in a dinucleotide repeat polymorphism in the tau gene between Caucasian and Japanese populations: implication for progressive supranuclear palsy. *Neurosci Lett* **250**: 135–137.

Davies KE (1986). *Human genetic diseases. A practical approach*. IRL Press: Oxford.

Evans W, Fung HC, Steele J, Eerola J, Tienari P, Pittman A *et al.* (2004). The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci Lett* **369**: 183–185.

Feuk L, Carson AR, Scherer SW (2006). Structural variation in the human genome. *Nat Rev Genet* **7**: 85–97.

Fujino Y, Wang DS, Thomas N, Espinoza M, Davies P, Dickson DW (2005). Increased frequency of argyrophilic grain disease in Alzheimer disease with 4R tau-specific immunohistochemistry. *J Neuropathol Exp Neurol* **64**: 209–214.

Fung HC, Evans J, Evans W, Duckworth J, Pittman A, de Silva R *et al.* (2005). The architecture of the tau haplotype block in different ethnicities. *Neurosci Lett* **377**: 81–84.

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C *et al.* (2001). Origins and divergence of the Roma (gypsies). *Am J Hum Genet* **69**: 1314–1331.

Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK (2007). Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *Eur J Hum Genet* **15**: 302–312.

Hardy J, Pittman A, Myers A, Fung HC, de Silva R, Duckworth J (2006). Tangle diseases and the tau haplotypes. *Alzheimer Dis Assoc Disord* **20**: 60–62.

Hardy J, Pittman A, Myers A, Gwinn-Hardy K, Fung HC, de Silva R *et al.* (2005). Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. *Biochem Soc Trans* **33**: 582–585.

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH *et al.* (2001a). Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet* **9**: 97–104.

Kalaydjieva L, Gresham D, Calafell F (2001b). Genetic studies of the Roma (Gypsies): a review. *BMC Med Genet* **2**: 5.

Kalaydjieva L, Morar B, Chaix R, Tang H (2005). A newly discovered founder population: the Roma/Gypsies. *Bioessays* **27**: 1084–1094.

Laws SM, Friedrich P, Diehl-Schmid J, Muller J, Eisele T, Bauml J *et al.* (2007). Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease. *Mol Psychiatry* **12**: 510–517.

Morar B, Gresham D, Angelicheva D, Tournev I, Gooding R, Guergueltcheva V *et al.* (2004). Mutation history of the roma/gypsies. *Am J Hum Genet* **75**: 596–609.

Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC *et al.* (2005). The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Hum Mol Genet* **14**: 2399–2404.

Myers AJ, Pittman AM, Zhao AS, Rohrer K, Kaleem M, Marlowe L *et al.* (2007). The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. *Neurobiol Dis* **25**: 561–570.

O'Hara RB (2007). Human expression patterns: genetic differences between populations. *Heredity* **98**: 245–246.

Pittman AM, Myers AJ, Duckworth J, Bryden L, Hanson M, Abou-Sleiman P *et al.* (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum Mol Genet* **13**: 1267–1274.

R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Semino O, Passarino G, Quintana-Murci L, Liu A, Beres J, Czeizel A *et al.* (2000). 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**: 339–346.

Sepkowitz KA (2006). Health of the world's Roma population. *Lancet* **367**: 1707–1708.

Skipper L, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M *et al.* (2004). Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet* **75**: 669–677.

Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG (2007). Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* **39**: 226–231.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J *et al.* (2005). A common inversion under selection in Europeans. *Nat Genet* **37**: 129–137.

Sundar PD, Yu CE, Sieh W, Steinbart E, Garruto RM, Oyanagi K *et al.* (2007). Two sites in the MAPT region confer genetic risk for Guam ALS/PDC and dementia. *Hum Mol Genet* **16**: 295–306.

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006). A map of recent positive selection in the human genome. *PLoS Biol* **4**: e72.

Warnes G, with contributions from G Gorjanc, F Leisch and Michael Man (2008). Genetics: Population Genetics, R package version 1.3.3.

Winkler S, Konig IR, Lohmann-Hedrich K, Vieregge P, Kostic V, Klein C (2007). Role of ethnicity on the association of MAPT H1 haplotypes and subhaplotypes in Parkinson's disease. *Eur J Hum Genet* **15**: 1163–1168.