



ARTICLE

Y chromosome analysis reveals a sharp genetic boundary in the Carpathian region

Mihaela Stefan¹, Gheorge Stefanescu², Lucian Gavrila¹, Luciano Terrenato³, Mark A Jobling⁴, Patrizia Malaspina³ and Andrea Novelletto^{3,5}

¹Genetics Department, University of Bucharest; ²Institutul Cercetari Biologice, Iasi, Romania; ³Department of Biology, University of Rome 'Tor Vergata', Italy; ⁴Department of Genetics, University of Leicester, UK; ⁵Department of Cell Biology, University of Calabria, Italy

Nine single nucleotide (SNP) or indel binary polymorphisms were used to determine the frequencies and phylogenetic relationships of 12 Y chromosomal haplogroups in 289 males from Romania and the Republic of Moldova. Our data indicated a low but not null rate of the homoplastic appearance of the DYZ3 (–) allelic state. All other markers confirmed the previously proposed phylogeny. Based on the affinities between populations in terms of haplogroup frequencies, this work identified the geographical region of the Carpathians as a break point in the gene geography of Eastern Central Europe, providing a finer definition of one of the possible sharp genetic changes between Western and Eastern Europe. *European Journal of Human Genetics* (2001) 9, 27–33.

Keywords: Y chromosome; population structuring; peopling of Europe

Introduction

The human Y chromosome is inherited only by males with transmission from fathers to sons, with the exception of the two pseudo-autosomal regions. After initial reports on the paucity of polymorphic DNA markers on this chromosome,^{1,2} extensive research resulted in a repertoire of loci useful in phylogenetic^{3–5} and population analysis. Haplotypes obtained by typing multiple Y-specific loci have proved to be highly valuable in describing affinities between populations in terms of this portion of the genome. Furthermore, a comparison between this male-borne portion of genetic diversity with that of the female-borne mtDNA, becomes even more promising for the description and understanding of the relative contributions of the two sexes to processes leading to the overall variation in the gene pools of extant populations. These include differential mating patterns and behaviour following marriage.⁶

There are three relevant features regarding the molecular diversity of the Y chromosome:

- (a) the unequivocal reconstruction of its phylogeny as revealed by non-recurrent single nucleotide (SNP) or indel binary polymorphisms;^{5,7,8}
- (b) the abundance of microsatellite markers which reveal additional diversity within SNP-defined lineages (haplogroups);^{9–11} and
- (c) the effectiveness with which this chromosome is able to detect population structuring better than either autosomal or mitochondrial markers.^{12,13}

Through a combination of Y-chromosomal binary and microsatellite markers, Malaspina *et al*¹⁴ have shown peculiar geographic distributions of certain lineages across Europe, suggesting that Central Europe might be the area where some of these lineages undergo the sharpest changes in frequencies. Here we describe the frequencies of Y-chromosomal haplogroups as defined by nine SNP markers in eight groups of males sampled in different regions of Romania and in the Republic of Moldova.

Materials and methods

Subjects

We analysed 219 males from eight administrative districts of Romania (samples 1–8) and 70 males from the Republic of

Correspondence: Dr Patrizia Malaspina, Department of Biology, University of Rome 'Tor Vergata', Via della Ricerca Scientifica, 00133 Rome, Italy. Tel: +39 06 72594321; Fax: +39 06 2023500; E-mail: patrizia.malaspina@uniroma2.it
Received 10 May 2000; revised 29 August 2000; accepted 13 September 2000

Moldova (sample 9) (Figure 1 and Table 1). The eight Romanian samples represented the country's three main geographical regions, ie Transylvania in the north-west (samples 1–5), Walachia in the south (sample 6) and Moldavia in the East (samples 7, 8) up to the right bank of the Prut river, setting the border with the Republic of Moldova (sample 9). Informed consent was obtained in all cases. The groups of Maramures (sample 1) and Vrancea (sample 8) were reported previously as Northern and Eastern Romania, respectively;¹⁴ subjects from these groups were typed with the additional markers here used.

DNA markers

The following binary polymorphisms were analysed by PCR followed by direct agarose or polyacrylamide gel fractionation or restriction enzyme digestion and agarose gel analysis: YAP insertion,¹⁵ DYS257,⁷ SRY₁₀₈₃₁^{4,7,16} (also known as SRY-1532), the presence/absence of the aliphoid *HindIII* site (DYZ3),¹⁷ M9,⁵ p12f2¹⁸ (and Rosser *et al*, in preparation), LLY22g/*HindIII*,¹⁹ Tat,¹⁹ 92R7.²⁰

The phylogenetic relationships proposed for haplogroups defined through the above markers by several authors can be summarised as shown in Figure 2: haplogroup 2 is ancestral to

haplogroups YAP(+) [insertion of the YAP element], 9 [rearrangement at locus p12f2] and 26 [M9(C) (G)]; haplogroup 26 is ancestral to haplogroups 1 [DYZ3(+) (–), DYS257(G) (A) and 92R7(C) (T)] and 12 [LLY22g(C) (A)]; haplogroup 1 is ancestral to haplogroup 3 [SRY₁₀₈₃₁(G) (A)] and haplogroup 12 is ancestral to haplogroup 16 [Tat(T) (C)]. Haplogroup 7, ancestral to haplogroup 2 and characterised by SRY₁₀₈₃₁(A) as an ancestral state, is virtually absent in all populations outside sub-Saharan Africa^{7,16,21} and was not searched in this work. We then adopted the following typing scheme to determine haplogroup frequencies in the nine population samples here examined. DYZ3, YAP and DYS257 were typed in all subjects. SRY₁₀₈₃₁ was typed on all DYS257(A) and a subset of DYS257(G). M9 was typed on all YAP(–)/DYS257(G) and a subset of DYS257(A). p12f2 was typed on all YAP(–)/M9(C) in addition to a subset of M9(G). LLY22g was typed on all DYS257(G)/M9(G) and other subjects. Tat was typed on all LLY22g(A) and other subjects. 92R7 was typed on 48 subjects only. By checking marker alleles in subsets of chromosomes where no variation was expected, the results were still consistent with the phylogeny outlined above. These alleles are identified by superscripts in Table 1 and the number of tested subjects is listed in the corresponding footnote.

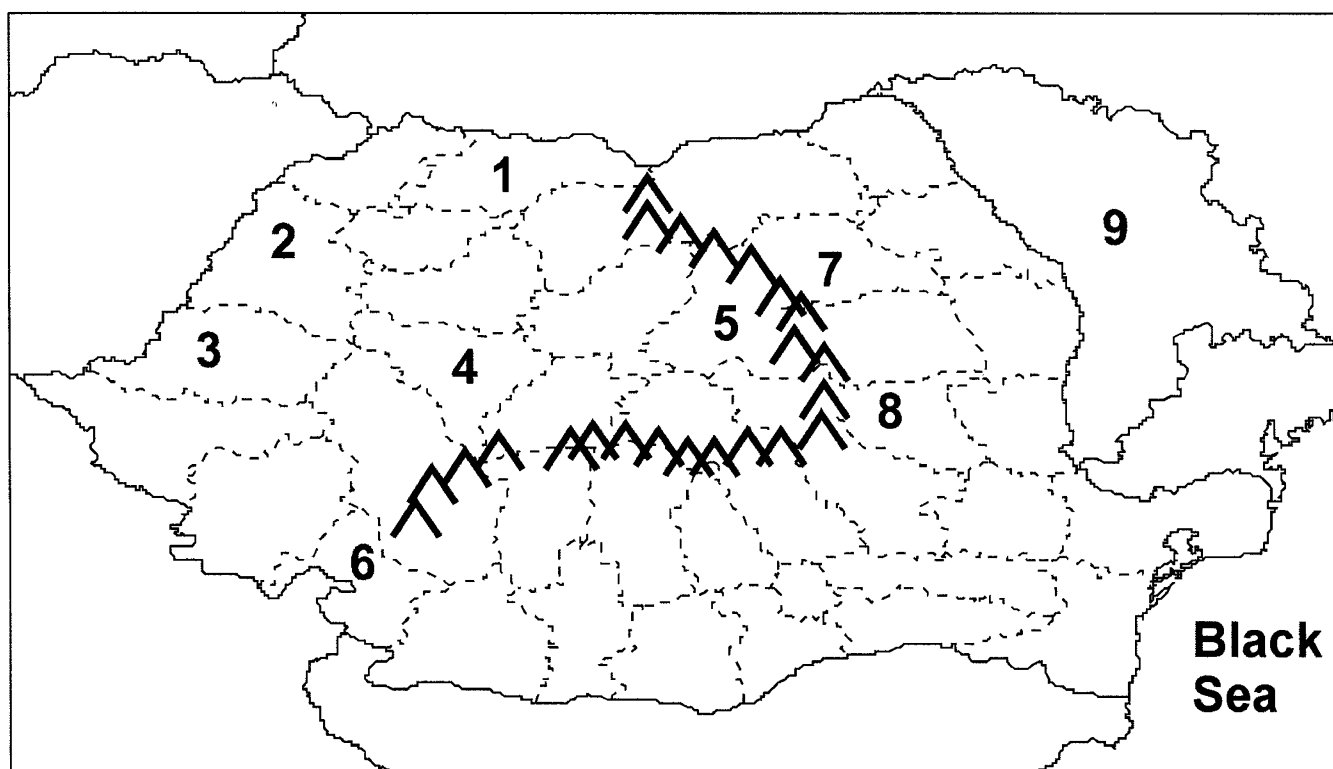


Figure 1 Map of Romania and the Republic of Moldova and location of the nine population samples. National borders are represented as solid lines. Borders of administrative districts are reported as broken lines. The position of the Carpathian watershed is highlighted. The Romanian geographical regions of Transylvania, Walachia and Moldavia are represented by samples 1–5, 6 and 7, 8, respectively.

Table 1 Structure, and absolute and relative frequencies of Y- chromosomal haplogroups in the nine population samples examined

No.	DYS 257	DYZ 3	YAP	Haplogroup ^b					Samples									Total	
				SRY 10831	p12 f2	M9	Lly 22g	Tat	92 r7	Romania (administrative district) ^a									
										1	2	3	4	5	6	7	8	R.M. 9	
1	A	-	-	G	+ ⁴	G ⁷	C ¹¹	T ¹⁶	T ²¹	7	5	9	8	9	3		2	4	47
										25.0	14.3	32.1	30.8	25.0	20.0		11.1	5.7	16.3
2	G	+	-	G ¹	+	C	C ¹²	T ¹⁷	C ²²	6	12	5	8	16	5	14	6	26	98
										21.4	34.3	17.9	30.8	44.4	33.3	42.4	33.3	37.1	33.9
2v	G	-	-	G	+	C	C	T	C	1									1
										3.6									0.3
3	A	-	-	A	+ ⁵	G ⁸	C ¹³	T ¹⁸	T ²³	4	7	4	5	7	1	9	5	25	67
										14.3	20.0	14.3	19.2	19.4	6.7	27.3	27.8	35.7	23.2
YAP ⁺	G	+	+	G ²	+ ⁶	C ⁹	C ¹⁴	T ¹⁹	C ²⁴	1	2	1	3	2	3	2	2	6	22
										3.6	5.7	3.6	11.5	5.6	20.0	6.1	11.1	8.6	7.6
YAP ^v	G	-	+			C												1	1
																		1.4	0.3
9	G	+	-	G ³	-	C ¹⁰	C ¹⁵	T ²⁰	C ²⁵	9	6	5			3	7	3	2	35
										32.1	17.1	17.9			20.0	21.2	16.7	2.9	12.1
9v	G	-	-	G	-	C		T	C										1
																			3.6
16	G	+	-		+	G	A	C			2	1				1		3	7
											5.7	3.6				3.0		4.3	2.4
16v	G	-	-	G	+	G	A	C	C					1				1	2
														2.8				1.4	0.7
26	G	+	-		+	G	C	T			1	2	2	1				1	7
											2.9	7.1	7.7	2.8				1.4	2.4
26v	G	-	-	G	+	G	C	T	C									1	1
																		1.4	0.3
Total										28	35	28	26	36	15	33	18	70	289

^a1: Maramures, 2: Bihor, 3: Arad, 4: Alba, 5: Harghita, 6: Mehedinti, 7: Neamt, 8: Vrancea, 9: Republic of Moldova (see also Figure 1);
^bAlleles not critical for defining haplotypes and for which no variation is expected were checked on some chromosomes only and are identified by superscripts: 1: 2 tested only; 2: 1; 3: 2; 4: 21; 5: 19; 6: 18; 7: 15; 8: 26; 9: 16; 10: 18; 11: 10; 12: 15; 13: 10; 14: 3; 15: 13; 16: 16; 17: 17; 18: 13; 19: 4; 20: 14; 21: 8; 22: 12; 23: 9; 24: 3; 25: 12. 'v': variant haplogroups characterised by DYZ3 (-).

Data analysis

A general overview of the data was obtained by generating a variety of genetic distance measures with the TFPGA²² and Arlequin 1.1²³ packages, and the corresponding dendrograms with the UPGMA algorithm. Analysis of molecular variance (AMOVA²⁴), was performed with the program Arlequin 1.1. In this method, a hierarchical analysis of variance partitions the total variance into covariance components due to inter-individual differences and inter-population differences. In the analysis, the number of mutations between molecular haplotypes may also be taken into account. In both cases, the total variance is the sum of the covariance component due to differences among haplotypes within a population, the covariance component due to differences among haplotypes in different populations within a group and the covariance component due to differences between the populations. The covariance components are used to compute fixation indexes (indicated by the letters Φ or F depending on whether molecular information is included, or not), ie F_{st} or the correlation of random haplotypes within populations relative to that of random pairs drawn from the entire study, F_{ct}

or the correlation of random haplotypes within a group of populations relative to that of random pairs of haplotypes drawn from the whole study, and F_{sc} or the correlation of random haplotypes within a population relative to that of random pairs of haplotypes drawn from the group of populations. The significance of the fixation indexes is tested using permutations to generate null distributions.

Haplotypes carrying the variant *HindIII*(-) at DYZ3 were pooled with the corresponding major haplogroups carrying (+) (see Results). Matrix entries of inter-haplotype distances to obtain Φ statistics corresponded to the numbers of mutational steps separating each pair of haplogroups, disregarding both 92R7 and DYZ3. Significance of F or Φ statistics was obtained with 1000 permutations.

Results

Out of a total of 12 haplogroups observed (Table 1), only four reached a frequency of >0.10 in the overall sample. Although our typing scheme is based on a precise underlying phylogenetic tree, DYZ3 and DYS257 were typed in all

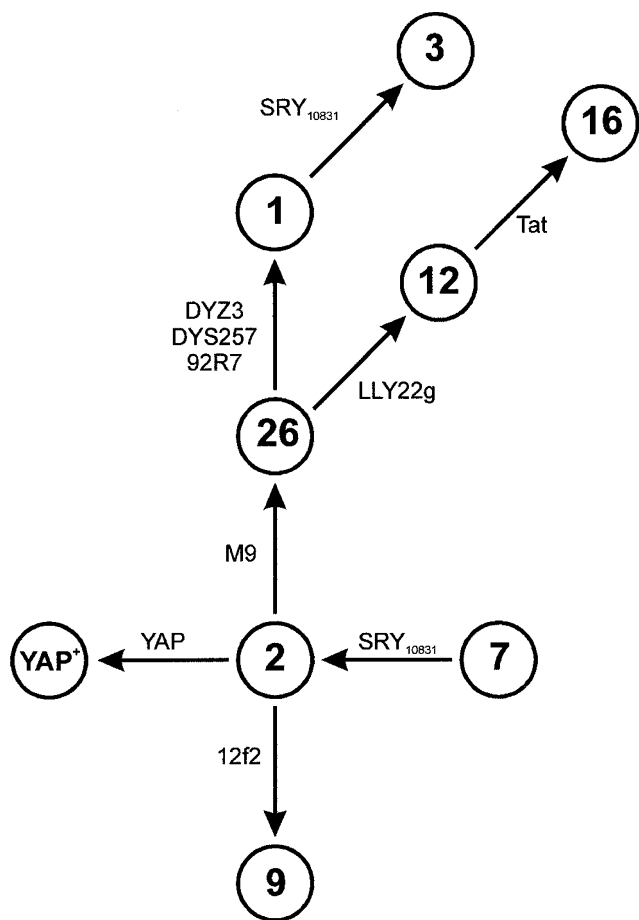


Figure 2 Schematic representation of the phylogenetic relationships among the Y-chromosomal haplogroups here detected. Each haplogroup is represented as a circle with the corresponding designation within it. The markers that define each mutational step are reported. Arrows are oriented from the ancestral to the derived state of each marker (modified from Bosch *et al*¹⁰).

subjects, in an attempt to infer the relative order of the corresponding mutational events. Out of 289 subjects, 283 carried either DYS257(G)/DYZ3(+) or DYS257(A)/DYZ3(-), showing that the two allele states at both loci largely mark the same branch of the tree leading from haplogroup 26 to haplogroup 1 and its derivative 3. In the six remaining subjects, the DYZ3(-) state was found in association with DYS257(G) on five different haplogroup backgrounds. In the case of 2v, this background could be distinguished from that of haplogroup 1 given the presence of M9(C). In three cases (YAP + v, 9v and 16v) one or more intervening mutations were incompatible with the DYZ3(-) allele state being shared with haplogroups 1 and 3 by descent. In only one case (26v) and DYZ3(-) state would have suggested this haplotype as a candidate ancestor of haplogroups 1 and 3. In all cases the *HindIII*(-) variant haplogroups carried 92R7(C), an allele usually associated with DYS257(G). In conclusion, the five

variant haplogroups can be accounted for mostly by the homoplastic appearance of the DYZ3(-) allele on several branches stemming out from haplotype 2. Such pattern was already evidenced on chromosomes carrying the YAP element or the ancestral SRY₁₀₈₃₁(A) allele.^{17,25} For this reason haplogroups carrying DYS257(G)/DYZ3(-) were pooled with the corresponding major haplogroups to analyse population diversity, whilst the DYZ3 *HindIII* site was not considered in the AMOVA analysis with molecular distances.

Geographic distributions and population heterogeneity

The overall heterogeneity of haplotype frequencies in the nine population samples was significant as assayed by chi-square ($\chi^2 = 80.22$, 48 d.f., $P < 0.005$). When haplogroup frequencies were plotted on the corresponding locations (Figure 1), at least three haplogroups showed coherent patterns of variation. Haplogroup 1 had the lowest frequencies (0–11.1%) in the three easternmost areas (samples 7–9), and frequencies of 20–32.1% in samples 1 and 3–6. For this haplogroup the Bihor district (sample 2) was rather distinct from the neighbouring areas, with an intermediate value of 14.3%. Haplogroup 3 had the highest frequencies (27.3–35.7%) in the easternmost regions (samples 7–9), nearly double those found in all other regions (6.7–20%). Haplogroup 9 was absent in the mountain areas of Central Transylvania (samples 4 and 5).

Genetic distance UPGMA trees obtained with all measures except Reynolds' co-ancestry coefficient clustered the nine samples into three major clades. The first included samples 1, 3 and 6, the second samples 4 and 5, and the third samples 2, 7, 8 and 9. Figure 3 displays the tree obtained with Slatkin corrected F_{st} .²⁶

These patterns prompted us to assay population structuring in relation to the main feature of the region investigated, ie the geographical location of each sample with respect to the Carpathian ridge. After carrying out a hierarchical grouping of population samples into two major groups, the different quotas of genetic diversity were evaluated by (a) multiallelic and (b) haplotypic AMOVA;²⁴ all haplogroups were considered equidistant in (a) whilst information on the number of mutational steps discriminating each pair of haplogroups was taken into account in (b). The corresponding overall F_{st} and Φ_{st} statistics were 0.027 ($P = 0.006$) and 0.013 (n.s.), respectively. These figures indicate that the vast majority of genetic variance is among individuals within populations. The residual portion thus measures a marginal effect in population structuring, which may be nevertheless relevant to model population processes in the past. The difference in the absolute values and significance levels of F_{st} and Φ_{st} statistics is not surprising as most of the structuring is contributed by haplogroups 1 and 3 (accounting for 50% of subjects). Since these haplogroups differ by a single mutational step out of the seven here considered, the inclusion of molecular information only adds more noise than signal to the Φ statistics. This can be interpreted by considering that

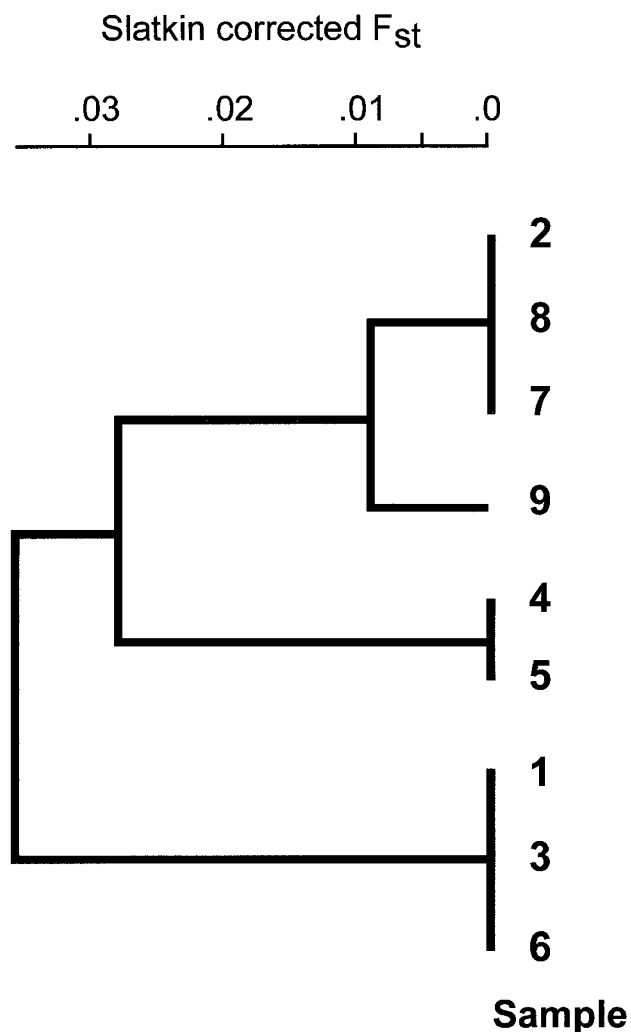


Figure 3 UPGMA tree based on linearised F_{st}^{30} for pairwise comparisons between the nine population samples considered.

all the haplogroups detected in this study arose over a long timespan outside the area here considered. It was only later that these haplogroups appeared in this area through migration. Therefore, the demographic processes presumably leading to the observed pattern of genetic diversity seem to have occurred on a shorter time scale than the accumulation of mutations on the chromosomes, as is the rule over much of the world and for most markers.²⁷⁻³⁰ As a consequence, the differences among populations are more faithfully described by differences in allele frequencies than by measures of molecular distance between haplogroups or chromosomes (Table 2).

An initial grouping of populations into a cluster east of the Carpathian watershed resulted in a F_{ct} of 0.034, the second highest value found in the distribution of all the 84 permutations of three vs six populations ($P < 0.012$). Correspondingly, F_{sc} was only 0.008 ($P = 0.19$), showing that most of the population structuring could be accounted for by differences

Table 2 Parameters of population structuring according to different grouping of samples (numbered as in Table 1 and Figure 1)

Samples grouped	F_{ct}	F_{sc}
1, 3, 6 vs others	0.043 ($P < 1/84$)	0.008
1, 2, 3, 4, 5, 6 vs others	0.034 ($P = 1/84$)	0.008
1, 3, 4, 6 vs others	0.037 ($P < 1/126$)	0.008
1, 3, 4, 5, 6 vs others	0.032 ($P = 1/126$)	0.009
1, 2, 3, 4, 6 vs others	0.031 ($P = 2/126$)	0.009
1, 3, 6 vs 4, 5 vs 2, 7, 8, 9	0.046 ($P < 1/1023$)	-0.005

between the two geographically defined groups. Alternative schemes for grouping the samples into two clusters were systematically searched. The five highest F_{ct} were found when one of the groups included the western samples 1, 3 and 6 and the other included the eastern samples 7, 8 and 9.

When three-group schemes were searched, the highest F_{ct} was obtained when clusters including samples 1, 3 and 6, 4 and 5, and 2, 7, 8 and 9 were compared (Table 2, last line).

Discussion

When considering the combination of marker alleles into haplogroups, our data indicate a low, yet not null, rate of the homoplastic appearance of the DYZ3(-) allelic state. While a major branch of the tree connecting the haplogroups here found is characterised by the DYZ3(-) state, in few instances (4%) it recurs on other backgrounds. This phenomenon is confusing when searching for intermediates between haplogroups 26 and 1, a step also marked by a change at DYS257 and 92R7 allele states. We found a single instance of a haplotype (26v) candidate to be a representative of such an intermediate. An additional non-recurrent marker of a subset of M9(G) haplogroups would be necessary to reinforce the idea that 26v is indeed ancestral to haplogroup 1.

On the other hand, the uncommon occurrence of DYZ3(-) as a variant of some haplogroups can be considered a useful tool in population studies. In fact, it may mark local subsets of chromosomes. The observation of two subjects with haplogroup 16v among nine Tat(C) subjects may be one such case, and may prompt the typing of other Tat(C) sets for DYZ3.^{19,31}

Our analysis shows that the Carpathians coincide with a genetic boundary within eastern Central Europe, where the area east of the Carpathians is characterised by frequencies one third (for haplogroup 1) and double (for haplogroup 3) with respect to those West of the Carpathian ridge. The only exception is represented by the Bihor sample (sample 2), which is located at the Hungarian border but clusters with the eastern samples in some of the analyses.

We suggest that the association between the mountain ridge and this genetic discontinuity is not a mere coincidence, since the Carpathians are also representative of an ecological boundary. In fact they constitute the south-western limit of a vast steppe area also delimited by the

northern coasts of the Black Sea. In prehistory a large population may have lived in this area (for a review see Levine³²), leaving a relevant contribution to the present-day gene pool. Cavalli-Sforza *et al*³³ showed that extreme values of the third synthetic variable summarising modern day distributions of polymorphic markers are centred in the same area. The main Y-chromosomal correlate of the above distribution is represented by the SRY₁₀₈₃₁(A) allele. According to Zerjal *et al*,⁸ in Europe SRY₁₀₈₃₁(A) reaches frequencies > 20% north of the area investigated in the present study and in Germany. These authors suggest that haplogroup 3 arose in Asia and subsequently spread to Europe. Thus, the Y-specific population-structuring here described may testify in favour of the poor overlap between the peopling of steppe and plain vs mountain environments, the latter being in closer continuity with the rest of Western Europe. The ecological discontinuity may have acted as a barrier to gene flow over long periods of time, allowing a divergence whose signature is still detectable in the population samples examined in this work. In view of their likely neutrality, Y-chromosomal markers are to be taken solely as indicators of such a divergence, no matter if it was attained by drift, selection, or both.

Another marker establishing connections with geographical areas north-east of the one here examined is Tat(C), which characterises haplogroup 16. We found this marker in the Republic of Moldova. Zerjal *et al*¹⁹ dated this mutation at < 4000 ybp and located it in Mongolia. It is interesting to observe that for both haplogroups 16 and 3, the immediate ancestral haplogroups (1 and 12, respectively), were detected at lower frequencies here, and their origins are also set in Asia.^{17,19} Thus, the presence of the Tat(C) allele too is also to be considered as an entry from a subset of the diversity found in Asia. Given the low estimate for the antiquity of this allele, this can be considered an independent and more recent event than that of SRY₁₀₈₃₁(A), possibly associated with the repeated invasions from the East during the first millennium AD.

Finally, chromosomes carrying p12f2(-) (also detectable as an 8 kb fragment on *TaqI* blots), characterising haplogroup 9, have been proposed to reflect the demic expansion associated with the spread of agriculture.³⁴ Our data reveal large variations ($P < 0.00015$) for the occurrence of this marker, with minimal frequencies as a peculiarity of samples from Central Transylvania. If the 'wave of advance' model for its spread is correct, one must conclude that Central Transylvania was poorly accessible to this chromosome, which flowed from the south-east to both east and west of the Carpathians. This hypothesis is compatible with the data for Greece and Albania.³³ Moreover, the recently reported³⁵ frequency of this marker in Hungary (2.2%) has been attributed to a marginal effect of neolithic migrations in Eastern Europe and to a much later migration of northern and Uralic-speaking peoples in the Hungarian plain in the ninth century AD. This event may also be responsible for a

boundary detected in the analysis of 63 autosomal protein variant allele frequencies in Europe.³⁶

In conclusion, this work has identified the geographical region of the Carpathians as a break point in the gene geography of eastern Central Europe, by providing a finer definition of one of the possible sharp genetic changes in terms of the frequencies of the haplogroups observed in Western vs Eastern Europe.

Acknowledgements

We thank Dr M Lo Ponte for reviewing the manuscript. Work supported by PRIN MURST 1999 (LT), and CNR grants 97.00712.PF36 (LT), 99.02543.CT04 (AN). MAJ is a Wellcome Trust Senior Fellow, grant 057559.

References

- Jackubiczka S, Arnemann J, Cooke HJ, Krawczak M, Schmidtke J: A search for restriction fragment length polymorphism on the human Y chromosome. *Hum Genet* 1989; **84**: 86–88.
- Malaspina P, Persichetti F, Novelletto A *et al*: The human Y chromosome shows a reduced level of DNA polymorphism. *Ann Hum Genet* 1990; **54**: 297–305.
- Hammer MF: A recent common ancestry for Y chromosomes. *Nature* 1995; **378**: 376–378.
- Whitfield LS, Sulston JE, Goodfellow PN: Sequence variation of the human Y chromosome. *Nature* 1995; **378**: 379–380.
- Underhill PA, Jin L, Lin A *et al*: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; **7**: 996–1005.
- Jobling MA, Tyler-Smith C: Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 1995; **11**: 449–456.
- Hammer MF, Karafet T, Rasanayagam A *et al*: Out of Africa and back again: nested clastic analysis of human Y chromosome variation. *Mol Biol Evol* 1998; **15**: 427–441.
- Zerjal T, Pandya A, Santos FR *et al*: The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. In: Pahiha SS, Deka R and Chakraborty R (eds). *Genomic Diversity: Applications in Human Population Genetics*. Kluwer Academic/Plenum Publishers: New York 1999, pp 91–102.
- Kayser M, de Knijff P, Dieltjes P *et al*: Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 1997; **18**: 1602–1607.
- Bosch E, Calafell F, Santos FR *et al*: Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am J Hum Genet* 1999; **65**: 1623–1638.
- Mohyuddin AQ, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C: Identification and characterization of novel human Y-chromosome microsatellites from sequence database information. *Nucleic Acids Res* 2000; **28**: e8.
- Scozzari R, Cruciani F, Malaspina P *et al*: Differential structuring of human populations for homologous X and Y microsatellite loci. *Am J Hum Genet* 1997; **61**: 719–733.
- Seielstad MT, Minch E, Cavalli-Sforza LL: Genetic evidence for a higher female migration rate in humans. *Nat Genet* 1998; **20**: 278–280.
- Malaspina P, Cruciani F, Ciminelli BM *et al*: Network analyses of Y-chromosomal types in Europe, Northern Africa, and Western Asia reveal specific patterns of geographical distribution. *Am J Hum Genet* 1998; **63**: 847–860.
- Hammer MF, Horai S: Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 1995; **56**: 951–962.

- 16 Santos FR, Pandya A, Tyler-Smith C *et al*: The central Siberian origin for Native American Y chromosomes. *Am J Hum Genet* 1999; **64**: 619–628.
- 17 Santos FR, Pena SDJ, Tyler-Smith C: PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. *Gene* 1995; **165**: 191–198.
- 18 Casanova M, Leroy P, Boucekkine C *et al*: A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 1985; **230**: 1403–1406.
- 19 Zerjal T, Dashnyam B, Pandya A *et al*: Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 1997; **60**: 1174–1183.
- 20 Hurles ME, Veitia R, Arroyo E *et al*: Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* 1999; **65**: 1437–1448.
- 21 Karafet TM, Zegura SL, Posukh O *et al*: Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 1999; **64**: 817–831.
- 22 Miller MP: *Tools for Population Genetic Analyses* (TFPGA) 1.3. 1997. A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.
- 23 Schneider S, Kueffer J-M, Roessli D, Excoffier L: Arlequin vers. 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland, 1997.
- 24 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
- 25 Scozzari R, Cruciani F, Santolamazza P *et al*: Combined use of biallelic and microsatellite Y chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 1999; **65**: 829–846.
- 26 Slatkin M: Inbreeding coefficients and coalescence times. *Genet Res* 1991; **58**: 167–175.
- 27 Hammer MF, Spurdle AB, Karafet T *et al*: The geographic distribution of human Y chromosome variation. *Genetics* 1997; **145**: 787–805.
- 28 Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G: Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci USA* 1998; **95**: 9053–9058.
- 29 Stoneking M, Fontius J, Clifford SL *et al*: Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 1997; **7**: 1061–1071.
- 30 Excoffier L, Schneider S: Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl Acad Sci USA* 1999; **96**: 10597–10602.
- 31 Carvalho-Silva DR, Santos FR, Hutz MH, Salzano FM, Pena SDJ: Divergent human Y-chromosome microsatellite evolution rates. *J Mol Evol* 1999; **49**: 204–214.
- 32 Levine M (ed): *Proceedings of the Symposium 'Late prehistoric exploitation of the Eurasian Steppe'*. McDonald Institute for Archeological Research; Cambridge, UK, 2000.
- 33 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton University Press: Princeton, NJ, 1994.
- 34 Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benecetti AS: A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 1996; **59**: 964–968.
- 35 Semino O, Passarino G, Quintana-Murci L *et al*: MtDNA and Y chromosome polymorphisms in Hungary: inferences from palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. *Eur J Hum Genet* 2000; **8**: 339–346.
- 36 Barbujani G, Sokal RR: Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA* 1990; **87**: 1816–1819.