International Congress Series 1261 (2004) 344–346

www.ics-elsevier.com

Y chromosome genetic structure in the Italian peninsula

C. Capelli*, B. Arredi, L. Baldassarri, I. Boschi, F. Brisighelli, A. Caglià, M. Dobosz, F. Scarnicci, V.L. Pascali

Forensic Genetics Laboratory, Istituto di Medicina Legale, Università Cattolica del S. Cuore, Largo F. Vito, 1-00168, Rome, Italy

Abstract. We investigated Y chromosome populations structure in the Italian peninsula by means of a number of microsatellites and SNPs markers. Genetic analysis revealed a certain degree of heterogeneity between samples at both haplotype (hpt) and haplogroup (hg) level. Implications in genetic history and forensic are discussed. © 2003 Published by Elsevier B.V.

Keywords: Y chromosome; SNP: single nucleotide polymorphism; STR: short tandem repeat; Population

1. Introduction

Y chromosome analysis has its own, well-established role in modern forensic DNA analysis. A vast consensus has been reached on the adoption of a 8-STRs haplotype (hpt), this being the base for the establishment of large national and continental databases. Y-STR databases have also been used to explore population genetic issues, either relevant to the forensic inference or of general concern. It is often stated that the distribution of Y hpts in large populations (such as national groups and big urban aggregates with homogenous ethnic composition) reflects the virtual absence of internal structure. We recently discussed the general validity of this assumption [1]. We in fact showed that significant differences actually exist between Central–North and Central–South Italian populations by the means of chromosome groups (haplogroups, hgs) defined by slowly evolving markers (Unique Event Polymorphisms, UEPs). This result is of central interest, in light of the fact that hpts and hgs are phylogenetically related to each other and haplotypes might mirror the same differential distribution shown by haplogroups. In order to investigate this specific issue in detail, we collected new samples from numerous local communities of Italy and carried out extensive Y-linked microsatellites and UEPs analysis.

* Corresponding author. Tel.: +39-6-35507031; fax: +39-6-35507033.
E-mail address: c.capelli.@rm.unicatt.it (C. Capelli).

0531-5131/ © 2003 Published by Elsevier B.V.
doi:10.1016/S0531-5131(03)01789-8

2. Sampling

Individuals were selected according to their father's father place of origin, to be within the area of interest. We included two samples from Central–North Italy (A and B), two samples from South Italy (D and E), and a Central–East location (C).

3. Y chromosome markers

The following SNPs were selected for the investigation: M35, M89, 12f2, M172, M170, M9, 92R7, M173, M17 [2–4]. Additionally, nine Y linked microsatellites were genotyped: DYS 388, 393, 392, 19, 390, 391, 389 I, 389 II and 385. The number of individuals genotyped for STRs/UEPs were: A 39/32; B77/58; C 57/58; D60/61; E 60/71.

4. Distribution of genetic variation

We investigated the issue of differential distribution of diversity across the samples areas by AMOVA and principal components (PC) analyses. The two approaches focused on microsatellite and SNP variation, respectively. The percentage of variation observed at haplotype level was largely within populations (96.71%), as usually observed in human populations. Interestingly, 3.29% was instead present between populations. This pattern was found to be significant ($p < 0.01$) [5].

When focusing on SNPs, populations showed interesting differences in haplogroup frequencies. For comparison, we include populations from data available in literature [2,6]. In particular, two groups, namely, R1*xR1a1 and E3b were found to have inverse distribution along the peninsula. The former was the main type in the North, declining in frequency moving South, while the latter had its highest percentage in the South. HGs R1*xR1a1 and E3b, together with hg J2, were the three most common type along the peninsula (38%, 16% and 21%, respectively). Differently from the former two, J2 did not clearly show any pattern of differential distribution, with a range of frequency between 13% and 25%. In the PC plot, the northern samples (A and B) did tend to cluster with French and Basques, along axis 1, while C, D and E instead were together with other Mediterranean populations (Greeks and Calabrians). The main factors shaping the principal components plot were hg R1*xR1a1 and E3b (with loading factors of 0.443 and -0.338) for axis 1, while R1*xR1a1 and I were on axis 2 (0.167 and -0.35). Both A and B had the highest frequencies of R1*xR1a1 , around 50%, while C, D and E displayed the highest for E3b hg. The first two principal components summarised more than 62% of the total genetic variation (data not shown).

5. Differential distribution of haplotypes

The AMOVA results of significant amount of genetic variation across populations is raising the issue of nonrandom distribution of haplotypes across the different populations. Similarly, as haplotypes are evolutionary linked to the mutations defining haplogroups, the different frequencies of haplogroups in the analysed samples suggests that populations might have different frequencies for certain Y chromosome haplotypes. The haplotypes shared at least by two populations were eight. Only one type was shared across four populations, while none was present all across the five groups. Most of types were

uniquely present in each population. This is possible to be related to the high level of resolution offered by the nine genotyped microsatellites. The highest haplotype frequency was 5% ,observed in the A sample, and showing frequencies ranging between 0% and 1% in the others.

6. Conclusions

We here report that significant differences emerge among local groups both at hpt and hg level, making the reality of Y lineages stratification in Italy highly complex. In particular, at least two hgs show significant different frequencies between North and South samples, namely, R1*xR1a1 and E3b. Those two are the main factors shaping populations distribution in PC plot. In the plot, it appears that there is a certain tendency for independent clustering of the North and South Italian samples. We envision that the whole issue that could reflect different historical demographic phenomena occurred in the Mediterranean basin and in the Italian Peninsula during past millennia. These results also raise concerns on the appropriateness of the use of all-purpose ''Italian'' databases for forensic casework inference. The AMOVA analysis seems to support this, as the distribution of genetic variation was found to be significant across different groups ($p < 0.01$). Nonrandom distribution of Y haplotypes within national borders seems to be not unusual [4]. Our results, together with the observation in other countries, raise the issue of how appropriate would be the use of a single national database before exhaustively investigate the issue of population structure.

References

[1] A. Caglià, et al., in: B. Brinkmann, A. Carracedo (Eds.), Progress in Forensic Genetics 9, International Congress Series, vol. 1239, Elsevier, Amsterdam, 2003, pp. 307–310, January.
[2] O. Semino, et al., Science 290 (5494) (2000) 1155–1159.
[3] Y Chromosome Consortium, Genome Res. 12 (2) (2002) 339–348 (February).
[4] C. Capelli, et al., Curr. Biol. 13 (2003) 979–984.
[5] S. Schneider, et al., Genetics and Biometery Laboratory, University of Geneva, Geneva, 2000.
[6] E. Bosch, et al., Am. J. Hum. Genet. 68 (2001) 1019–1029.