ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International xxx (2004) xxx–xxx
Forensic Science International
www.elsevier.com/locate/forsciint

A collaborative study of the EDNAP group regarding Y-chromosome binary polymorphism analysis

María Brion[a,*], Berit M. Dupuy[b], Marielle Heinrich[c], Carsten Hohoff[c], Bernardette Hoste[d], Bertrand Ludes[e], Bente Mevag[b], Niels Morling[f], Harald Niederstätter[g], Walther Parson[g], Juan Sanchez[f], Klaus Bender[h], Nathalie Siebert[d], Catherine Thacker[i], Conceiçao Vide[j], Angel Carracedo[a]

[a]Institute of Legal Medicine, University of Santiago de Compostela, San Francisco s/n, 15782 Santiago de Compostela, Spain
[b]Institute of Legal Medicine, Oslo, Norway
[c]Institute of Legal Medicine, Münster, Germany
[d]National Institute of Forensic Science, Brussels, Belgium
[e]Institute of Legal Medicine, Strasbourg, France
[f]Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Denmark
[g]Institute of Legal Medicine, Innsbruck, Austria
[h]Institute of Legal Medicine, Mainz, Germany
[i]Institute of Cell and Molecular Science, Barts and The London, Queen Mary's School of Medicine and Dentistry, London, UK
[j]Institute of Legal Medicine, Coimbra, Portugal

Received 20 May 2004; accepted 10 June 2004

Abstract

A collaborative study was carried out by the European DNA Profiling Group (EDNAP) in order to evaluate the performance of Y-chromosome binary polymorphism analysis in different European laboratories. Four blood samples were sent to the laboratories, to be analysed for 11 Y-chromosome single nucleotide polymorphisms (SNPs): SRY-1532, M40, M35, M213, M9, 92R7, M17, P25, M18, M153 and M167. All the labs were also asked to submit a population study including these markers.

All participating laboratories reported the same results, indicating the reproducibility and robustness of Y-chromosome SNP typing.

A total of 535 samples from six different European populations were also analysed. In Galicia (NW Spain) and Belgium, the most frequent haplogroup was R1b*(xR1b1, R1b3df), which is almost absent in Austria and Germany. Haplogroup P*(xR1ab) is one of the most frequent in Austria, Germany and Norway and scarcely appears in Galicia, Belgium and Denmark.

Haplogroup frequencies found in this collaborative study were compared with previously published European Y-chromosome haplogroup data.
© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Y-chromosome; SNPs; DNA; EDNAP

1. Introduction

Analysis of Y-chromosome polymorphisms has already become a routine technique in most laboratories involved in forensic testing and kinship analysis. Although Y-STRs are the markers of preference in forensic labs, an increasing interest in Y-chromosome single nucleotide polymorphisms (SNPs) is evident in the field today [1].

SNPs are the simplest and most frequent kind of DNA sequence variation among individuals; their mutation rate is low and they can be analysed in short amplicons using new, high throughput technologies.

* Corresponding author. Tel.: +34-981582327; fax: +34-981580336.
E-mail address: brioniml@usc.es (M. Brion).

0379-0738/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2004.06.007

ARTICLE IN PRESS

Whether SNPs will replace STRs as primary method of choice in the forensic labs is a matter of conjecture at present. There is no doubt, however, about the usefulness of SNP typing for some specific purposes. For instance, Y-chromosome and mtDNA SNPs are informative in the analysis of geographical origin of a sample, which can be of importance in the investigation of criminal cases and for identification purposes.

As a consequence of the interest of several European labs in implementing analysis of Y-chromosome SNPs in routine work, a collaborative exercise was proposed by the European DNA Profiling Group (EDNAP). The first goal was to determine whether uniformity of results of Y-SNP typing could be achieved, independent of the methodology used for genotyping.

From the available described markers when the present exercise was planned, a set of 11 Y-chromosome binary polymorphisms was selected, to be typed for two control and two unknown individuals in each participating lab. Each lab was free to decide the methodology used for genotyping, according to what was available to them.

Previous to this exercise a preliminary trial was carried out with six Y-chromosome SNPs (M9, SRY-1532, SRY-8299, SRY-2627, 92R and TAT) with the aim of introducing SNP typing technologies in EDNAP labs.

The labs were also asked to perform a population study with the 11 markers. Y-chromosome haplogroup frequencies were reported from six different European populations and comparisons made between these populations and previously reported population studies.

2. Material and methods

2.1. Samples

Four blood stains from laboratory staff (Institute of Legal Medicine, Santiago de Compostela), with informed consent and previous acceptance of the local ethical committee, were sent to the 20 members of the EDNAP group. The samples included two control samples with known Y-SNP haplogroup profiles, and two unknown male samples.

In addition, six of the labs submitted population data. A total of 535 individuals were analysed, from Austria (Innsbruck area, 129), Belgium (54), Spain (Galicia, 100), Germany (Münster area, 95), Denmark (107) and Norway (51).

2.2. Genetic analysis

A total of 11 Y-chromosome binary markers were analysed: SRY-1532 (also known as SRY_{10831}), M40 (also known as SRY-8299 or SRY_{4064}), M35, M213, M9, 92R7, M17, P25, M18, M153 and M167 (also known as SRY-2627).

Table 1
Typical PCR conditions and restriction enzymes for SNP typing

SNP	PCR conditions		Restriction enzymes	Recognition sites
SRY-1532	Buffer	1×	*Dra*III	CACACA/GTG
M40			*Bsr*BI	GAG/CGG
M35	dNTPs	200 μM	*Bmr*I	ACTGGGACACA/G
M213	$MgCl_2$	1.5 mM	*Mae*II	A/CGT
M9			*Hinf*I	G/ANTC
92R7	Primers	0.25 μM each one	*Hind*III	A/AGCTT
M17	Taq	0.5 U		
P25	DNA	10 ng	*Cac*8I	GCN/NGC
M18	H_2O	Up 25 μl	*Mae*III	/GTTAC
M153			*Tsp*509I	/AATT
M167			*Bsi*HKAI	GTGCT/C

The participants were asked to type the samples for the 11 Y-SNPs using any technology that they had available. In order to facilitate the analysis, each group was provided with some information about each of the markers, including sequences of the described amplicons and primers in the literature (data not shown), PCR conditions and the restriction enzymes recognizing the polymorphic sites (Table 1).

The intention of the exercise was to see whether the typing results obtained were reproducible and accurate, independent of primers, PCR strategies and analytical techniques used.

Haplogroups were named according to the proposal of the Y-Chromosome Consortium [2,3].

2.3. Statistical analysis

Binary marker haplogroup frequencies were estimated by simple gene counting. The Arlequin software version 2.0 [4] was used to estimate haplogroup diversity values, to calculate genetic distances as pair-wise values of $\Phi_{ST,}$ and to perform analysis of molecular variance by means of AMOVA. Using the SPSS version 11.0 software package, principal component analysis (PCA) was performed on the haplogroup frequencies detected in the populations investigated and in previously studied populations [5–7].

3. Results

3.1. Inter-comparison exercise

Of the 20 member laboratories of the EDNAP group, 8 submitted results (previously 10 labs had sent results for the preliminary exercise with six Y-chromosome SNPs). All the submitted results were correct (Table 2). Several different technologies for genotyping had been used. Most of the labs used minisequencing, or single base extension method using

ARTICLE IN PRESS

Table 2
Results of the inter-laboratory exercise

Sample	SRY-1532	M40	M35	M213	M9	92R7	M17	P25	M18	M153	M167	HG
Control (1)	G	G	G	C	G	T	+G	C	−AA	T	C	P*(xR1ab)
Control (2)	G	G	G	C	C	C	+G	C	−AA	T	C	F*(xK)
Test (3)	G	G	G	C	C	C	+G	C	−AA	T	C	F*(xK)
Test (4)	G	G	G	C	G	T	+G	C	−AA	T	C	P*(xR1ab)

the SNaPshot multiplex kit (Applied Biosystems). One lab used direct sequencing and another lab used RFLPs and mass spectrometry.

Although PCR conditions for the described primers in the literature were provided to each lab, different primers and different PCR conditions (data not shown) were used. In all the cases, the results were completely concordant.

During the exercise, some labs independently reported double signals or non-consensus results for two markers, P25 and 92R7, and the existence of duplications in these markers was thus demonstrated. As a consequence, three EDNAP labs performed additional research on these duplicated markers [8].

3.2. Population data

In addition to the inter-lab comparison exercise, all the labs were asked to send population data on the Y-chromosome SNPs used in the study. A total of 535 individuals were analysed distributed among six populations, Austria (Innsbruck area), Belgium, Denmark, Germany (Münster area), Spain (Galicia) and Norway.

Although the set of Y-SNPs studied here allows the detection of 13 possible haplogroups, only 11 of them were observed in our populations (Fig. 1). The haplogroup composition of each population and haplogroup diversities are represented in Table 3. The highest diversity value was found in Norway, while the lowest one was observed in Galicia.

As shown in Fig. 1, the haplogroup composition of each population seemed to be quite different. To investigate this heterogeneity, an exact test of sample differentiation was undertaken using the haplogroup frequencies. In most cases significant deviations were found, except for Austria and Germany ($P = 0.166 \pm 0.022$; 10,000 Markov chain steps).

Haplogroup R1b*(xR1b1, R1b3df) was one of the most frequent in Galicia, Belgium and Denmark, while it was almost absent in Central Europe (Austria and Germany). Haplogroup F*(xK) is well represented in all the populations, while P*(xR1ab), which is one of the most frequent in

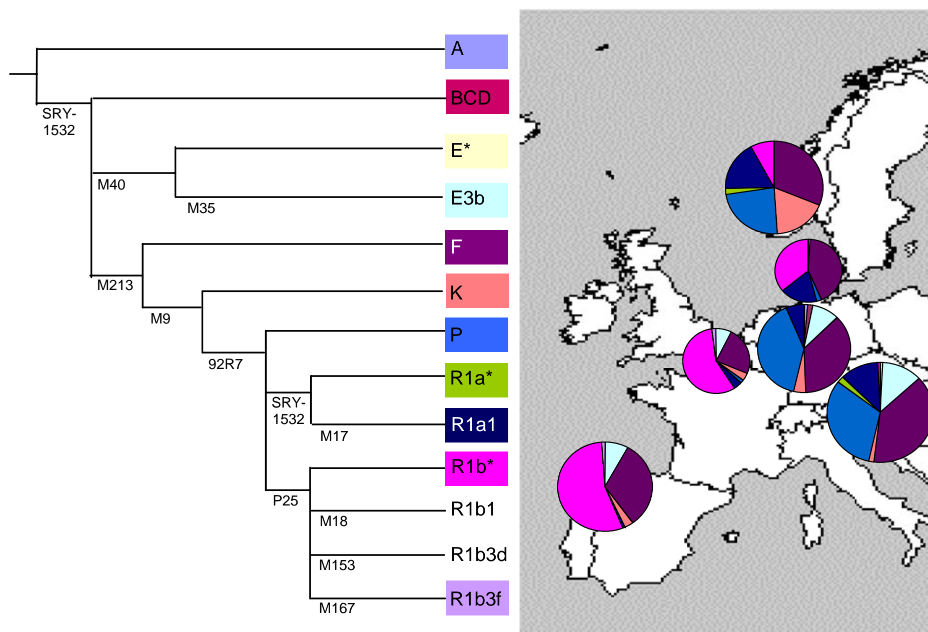

Fig. 1. (Left) Maximum parsimony phylogeny of the Y-chromosome SNPs analysed in the EDNAP exercise. Haplogroups found in the populations studied were represented with different colours. (Right) Relative haplogroup frequencies observed in each population.

ARTICLE IN PRESS

Table 3
Haplogroup composition and diversity values observed in the populations

Population	n	A*	BCD	E*(xE3b)	E3b	F*(xK)	K*(xP)	P*(xR1ab)	R1a*(xR1a1)	R1a1	R1b*(xR1b1, R1b3df)	R1b1	R1b3d	R1b3f	HgD[a]
Belgium	54				4	13	2	1		2	31			1	0.617 (0.060)
Denmark	107				1	45		3		19	38				0.664 (0.022)
Norway	51					16	9	12	1	9	4				0.793 (0.024)
Galicia	100				8	32	3			1	55			1	0.593 (0.035)
Innsbruck	129			1	16	50	2	41	3	15	1				0.725 (0.022)
Münster	95	1	2		9	35	4	38		6					0.696 (0.029)

[a] Haplogroup diversity (S.E. values).

Central Europe, scarcely appears in Belgium, Denmark and Galicia.

In order to compare the haplogroup composition detected in our collaborative exercise with previous European studies, a principal component analysis was performed using our data in conjunction with published data of European populations (Fig. 2). A total of 1491 samples distributed among 17 populations were plotted. Münster, Innsbruck and Norway clustered together, while Belgium and Galicia clustered with France, Italy and the Netherlands.

To assess the level of population structure, we estimated various Φ statistics by means of AMOVA [9]. The Φ_{st} value calculated for the entire European sample, comprising six populations without partitioning, was 0.085, indicating that a small proportion of the overall variation resulted from inter-population differences. When the populations were grouped according to geographical location, in North, Central and South Europe, a low degree of inter-group variability was observed (data not shown).

4. Discussion

In the forensic field, the use of inter-laboratory exercises has become a useful tool, especially as a first step of the validation procedure when new markers or technologies need to be implemented. There are several publications describing similar exercises; for example the reproducibility of Y-STR multiplexes, also carried out by the EDNAP group [10], and the quality assurance exercise performed by all the labs contributing to the Y-chromosome haplotype reference database (YHRD) [11]. These are only two examples, among others available in the literature.

The present exercise has demonstrated a clear uniformity in Y-chromosome SNP typing, independent of the strategy used. Among the eight labs collaborating in the control assay, four different technologies were used and all achieved the same results. Six labs chose single base extension using the SNaP-shot multiplex kit of Applied Biosystems, some of them as multiplex reactions and others as singleplex reactions.

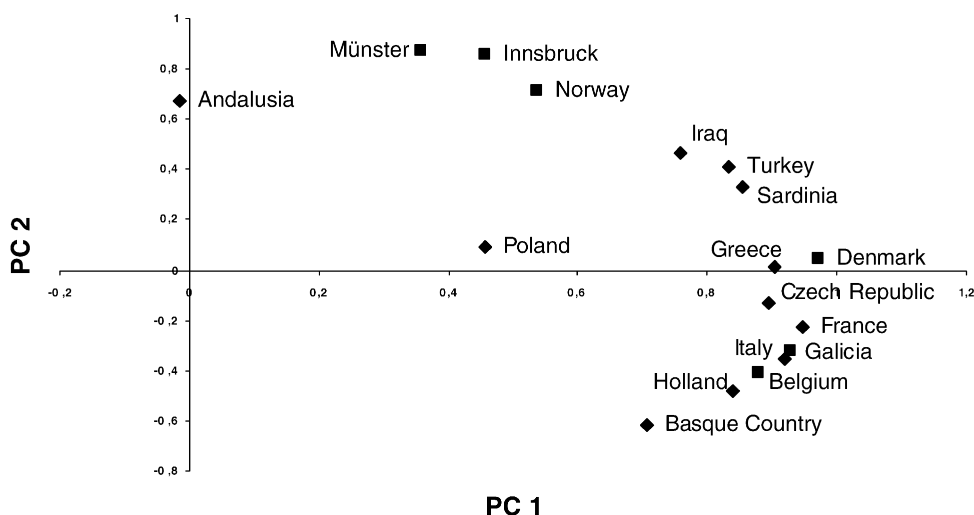

Fig. 2. Principal component (PC) analysis of haplogroup frequencies of several European populations. The first PC accounts for 58.7% of the variance, whereas the second accounts for 23.7%. Populations studied here are shown as squares and previously studied populations are shown as diamonds.

ARTICLE IN PRESS

The participating labs were asked to type a small sized population sample, in order to evaluate how informative the selected Y-chromosome SNPs were. Six of the laboratories submitted population data, and because all of them reported correct results in the collaborative exercise, their data was analysed. The population data from a seventh lab were complemented with Y-STR haplotype data and published independently [12].

It is important to avoid ascertainment bias in SNP selection for Y-chromosome studies [3]. In this case, among the 535 samples analysed, a total of 11 haplogroups were detected. Since the number of haplogroups described with the 11 SNPs included in the exercise is 13, it seems that the selected SNPs are appropriate for the populations analysed. When comparisons with previous European population studies were performed, our data showed no significant differences and clusters of neighbouring populations were seen in the PCA plot.

The distribution of Y-haplogroups showed significant differences between the population in the exercise. Although the differences were related to frequency, and not to the presence or absence of particular haplogroups in specific populations, the results show the potential usefulness of Y-chromosome SNP markers when inferring the geographical origins of samples.

Acknowledgements

The technical assistance of Meli Rodríguez and Raquel Calvo is highly appreciated. This work is part of the Ph.D. thesis of Marielle Heinrich.

References

[1] M. Brion, B. Sobrino, A. Blanco-Verea, M.V. Lareu, A. Carracedo, Hierarchical analysis of 30 Y-chromosome SNPs in European populations. Int. J. Legal Med. (2004) Apr 17.
[2] Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12 (2004) 339–348.
[3] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, Nat. Rev. Genet. 4 (8) (2003) 598–612.
[4] S. Schneider, D. Roessli, L. Excoffier, Arlequin ver. 2.000: A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000.
[5] O. Semino, G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova, L.E. Beckman, G. De Benedictis, P. Francalacci et al., The genetic legacy of paleolithic *Homo sapiens* sapiens in extant Europeans: a Y chromosome perspective, Science 290 (2000) 1155–1159.
[6] N. Al-Zahery, O. Semino, G. Benuzzi, C. Megri, G. Passarino, A. Torroni, A.S. Santachiara-Benerecetti, Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of early human dispersal and of post-Neolithic migrations, Mol. Phylogenet. Evol. 28 (2003) 458–472.
[7] C. Cinnioglu, R. King, T. Kivisild, E. Kalfoglu, S. Atasoy, G.L. Cavallery, A.S. Lillie, C.C. Roseman et al., Excavating Y-chromosome haplotype strata in Anatolia, Hum. Genet. 114 (2004) 127–148.
[8] J.J. Sanchez, M. Brion, W. Parson, A.J. Blanco-Verea, C. Boersting, M.V. Lareu, H. Niederstätter, H. Oberacher, N. Morling, A. Carracedo, Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications, Forensic Sci. Int. 140 (2004) 241–250.
[9] L. Excoffier, P.E. Smouse, J.M. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, Genetics 131 (1992) 479–491.
[10] A. Carracedo, A. Beckmann, A. Bengs, B. Brinkmann, A. Caglia, C. Capelli, P. Gill, L. Gusmao et al., Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19 DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format, Forensic Sci. Int. 119 (2001) 28–41.
[11] L. Roewer, M. Krawczak, S. Willuweit, M. Nagy, C. Alves, A. Amorim, K. Anslinger, C. Augustin et al., Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, Forensic Sci. Int. 118 (2-3) (2001) 106–113.
[12] K. Bender, B. Stradmann-Bellinghausen, C. Rittner, P.M. Schneider, Comparative analysis of short tandem repeats and single nucleotide polymorphisms on the Y-chromosome in Germans, Chinese and Thais, Legal Med. 5 (Suppl. 1) (2003) S164–S168.