

Short communication

Y-chromosome haplotypes in Italy: the GEFI collaborative database

S. Presciuttini^{a,*}, A. Caglià^b, M. Alù^c, A. Asmundo^d, L. Buscemi^e, L. Caenazzo^f,
E. Carnevali^g, E. Carra^h, Z. De Battistiⁱ, F. De Stefano^j, R. Domenici^k,
A. Piccinini^l, N. Resta^m, U. Ricciⁿ, V.L. Pascali^b

^aDepartment of Biomedicine, University of Pisa, S.S. Abetone e Brenero 2, 56127 Pisa, Italy

^bInstitute of Legal Medicine, Catholic University, Rome, Italy

^cDepartment of Morphological and Forensic Sciences, University of Modena, Modena, Italy

^dInstitute of Legal Medicine, University of Messina, Messina, Italy

^eChair of Legal Medicine, University of Ancona, Ancona, Italy

^fInstitute of Legal Medicine, University of Padova, Padova, Italy

^gChair of Legal Medicine at the Hospital of Terni, University of Perugia, Perugia, Italy

^hDepartment of Cellular Biology and Development, University of Palermo, Palermo, Italy

ⁱInstitute of Legal Medicine, University of Verona, Verona, Italy

^jInstitute of Legal Medicine, University of Cagliari, Cagliari, Italy

^kChair of Legal Medicine, Department of Biomedicine, University of Pisa, Pisa, Italy

^lInstitute of Legal Medicine, University of Milan, Milan, Italy

^mChair of Medical Genetics, Department of Internal Medicine, University of Bari, Bari, Italy

ⁿCenter of Medical and Molecular Genetics, Hospital "A. Meyer", Florence, Italy

Received 4 April 2001; accepted 6 April 2001

Abstract

A sample of 1176 males from 10 Italian regions have been typed for DYS19, DYS389-I, DYS389-II, DYS390, DYS391, DYS392, DYS393, and DYS385. Individual haplotype data are available on line. A low degree of variation is present among regions. Use of this database is specifically recommended for forensic applications in Italy. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Y-chromosome; Short tandem repeats; Population data

1. Population

Participating laboratories were asked to type at least 50 unrelated subjects born in their region for eight pre-defined STR loci. The organizing committee of the Forensic Geneticists of Italy (GEFI) group sent four blind control samples to each group (two single-profile stains and two admixtures, one male–male and one male–female, respectively). Data from the groups that returned blind-test correct results were included in the database. A total number of 1176 individuals were analyzed.

2. Extraction, PCR and typing

Laboratories were left free to use their preferred experimental procedures.

3. Typed loci

DYS19, DYS389-I, DYS389-II, DYS390, DYS391, DYS392, DYS393, DYS385. This locus ordering is used in the present paper to specify the allelic composition of given haplotypes. The haplotype at the first seven loci will be henceforward referred as to '7-L haplotype'; the entire profile (which is not a true haplotype, since the genotypes of

* Corresponding author.

DYS385 cannot be resolved into alleles of separate loci) as the ‘9-L haplotype’. Allele designation is consistent with the database of the International Forensic Y-User Group (IFYUG, http://ystr.charite.de/index_kl.html), where the number of repeats of DYS389-I and DYS389-II has been increased of three units compared to the nomenclature given in [1].

4. Data analysis

Data from different laboratories of the same region were pooled. The following labels are used throughout this paper: Lom, Lombardy (North); Ven, Veneto (North); Lig, Liguria (North); Em-R, Emilia-Romagna (North); Tus, Tuscany (Center); Mar, Marche (Center); Umb, Umbria (Center); Lat, Latium (Center); Pug, Puglia (South); Sic, Sicily (South). Calculations have been carried out separately for the 7-L haplotype, the DYS385 marker, and the combined 9-L haplotype.

5. Results

The 13 forensic laboratories contributed a total of 1176 males from 10 Italian regions, including 68.2% of the Italian population. Individual haplotype data are available at <http://www.gefi-forensicDNA.it>. Table 1 shows a summary of the diversity found in each region for the 7-L haplotype, in terms of the number of haplotypes that have been observed in one individual, in two individuals, and in n individuals (the frequency spectrum of the haplotype). A simple parameter related to haplotype diversity is the rate (H/N) of the total number of different haplotypes, H , to the total number of individuals in a sample, N (see last row of Table 1). The mean value of H/N across regions, weighted by sample size, was

0.78. In the global sample, we found a total of 650 different haplotypes in 1176 individuals ($H/N = 0.55$). The lower diversity in the total sample is a consequence of the non-infinite number of haplotypes effectively segregating in the general population. The proportion of unique haplotypes in the total sample was high (42%, in comparison with a sample weighted mean of 66.6%), meaning that the effective number of haplotypes in the population is still much larger than our total sample size. Table 2 shows the list of the 7-L haplotypes observed ≥ 5 times in the total sample, subdivided by region and ordered by decreasing total frequency. The most frequent haplotype in the total sample (14-13-29-24-11-13-13, $q = 0.057$) was the most frequent in each sub-sample also, with the exception of the two regions in the South. The same haplotype is the most frequent in Iberia [2] and in some samples from Center Europe, whereas it is almost absent in Eastern Europe (IFYUG on-line database). The entire list of allelic combinations observed at locus DYS385 is shown in Table 3. This locus is characterized by a relatively high frequency of a single genotype (11-14; $q = 0.21$ in the total sample) and a long tail of rare allele combinations, but it has a high discrimination capacity [3]. The match probability of two random individuals from the global sample for this locus alone was in our data 6.7%. The corresponding mean value among regions was 8.0%. Interestingly, the proportion of genotypes observed once in the total sample was small (1.0%), remarkably smaller than in sub-samples (weighted mean = 10.2%). This means that the size of our global sample approached for this locus the limit at which the genotypes effectively segregating in the population are all represented in the sample (the situation of sample saturation [4]).

Tables 3 and 4 shows a summary of the diversity found in each region and in the total sample for the entire 9-L profile. In one of the regions (Puglia) all haplotypes were unique ($H/N = 1$). In the other regions the weighted mean value of H/N

Table 1
L-7 haplotype frequency spectra in the 10 regions^a

No. of individuals per haplotype	North				Center				South	
	Lom	Ven	Lig	EmR	Tus	Mar	Lat	Umb	Pug	Sic
18					1					
10				1				1		
9						1				
6	1				1			2		
5					1					1
4		2			3			3		1
3	2	1	1	4	9	2	1	5	1	2
2	1	11	9	8	12	7	3	19	5	21
1	37	87	60	51	126	45	42	135	57	143
No. of different haplotypes (H)	41	101	70	64	153	55	46	165	63	168
No. of individuals (N)	51	120	81	89	218	74	51	222	70	200
H/N	0.80	0.84	0.86	0.72	0.70	0.74	0.90	0.74	0.90	0.84

^a The most common haplotype in the total sample is the most common in 8 out of 10 sub-samples.

Table 2

7-L haplotype frequency distribution in the 10 regions and in the total sample (haplotypes observed ≥ 5 in the total sample)^a

Haplotype	Lom	Ven	Lig	EmR	Tus	Mar	Umb	Lat	Pug	Sic	Total
14-13-29-24-11-13-13	6	4	3	10	18	9	3	10		4	67
13-13-30-24-10-11-13	3	2	1	2	2			6	3	5	24
14-13-29-24-10-13-13	1	2	2	1	5	3	2	4			20
14-13-29-23-11-13-13	1	4		1	3	2		3	1		15
14-13-29-23-10-11-12	2			1	2		1	6	1	2	15
14-14-30-24-11-13-13		2		3	4			4		1	14
14-13-30-24-11-13-13		1		1	4	1		1	1	2	11
14-13-29-23-10-13-13			1		4	2		2		1	10
15-13-29-23-09-11-12		1	2	1	2			2	1	1	10
15-13-29-23-10-11-12	1	1			3		1	1	1	2	10
15-12-29-22-10-11-13		1			2	1	1	3	1		9
15-12-29-22-10-11-14		1	2		1	2		2		1	9
14-13-29-24-11-14-13		1	1		2	1	1	3			9
14-13-29-24-11-12-13		3			6						9
14-13-29-25-11-13-13	1	1	1	3	3						9
14-13-29-24-11-13-12	1		1		3				2	1	8
15-13-29-24-11-13-13			1	3	3			1			8
14-13-29-24-10-13-12		2				2		2		1	7
15-12-28-22-10-11-14		2						1	2	2	7
14-13-29-24-10-12-13		2			3				2		7
14-13-28-24-11-13-13	3			1	3						7
14-12-28-23-10-11-13	1	1	1					4			7
14-12-28-24-11-13-13			2	2	1		1	1			7
15-12-30-22-10-11-14		1	1			2				2	6
14-13-30-24-10-11-13						1		2		3	6
15-12-28-24-10-11-12					3		2	1			6
14-13-30-23-10-13-13			1	1	1			1		2	6
14-13-30-23-10-11-12			2				1	2		1	6
14-13-29-23-10-11-13		1	1					3		1	6
13-12-29-24-10-11-13	1	1			1			1		2	6
15-12-29-23-10-11-14		2			2	1					5
13-13-29-24-10-11-13		1			1	1			1	1	5
13-13-31-24-10-11-13			1				1	1	1	1	5
14-13-29-23-11-11-12					2			1	1	1	5
14-13-30-22-10-11-12							1	2	1	1	5
13-13-30-25-10-11-13					2			1		2	5
14-14-31-23-10-11-12					3			2			5
13-13-30-23-10-11-13			2					1		2	5
Others	30	83	55	59	129	46	36	148	51	158	795
Total	51	120	81	89	218	74	51	222	70	200	1176

^a For locus ordering of the indicated haplotypes (first column) see Section 3.

Table 3

Frequency distribution of DYS385 allelic combinations in 1176 Italian males from 10 regions, ordered by decreasing total sample frequency

DYS385	Lom	Ven	Lig	EmR	Tus	Mar	Umb	Lat	Pug	Sic	Total
11-14	17	24	10	35	51	19	10	43	12	30	251
11-15	2	7	5	6	19	10	4	9	6	11	79
13-16	5	3	1		8	3	5	9	4	13	51
12-14	2	9	7	6	8	2	1	9		6	50
13-15	3	7	5	1	8		3	7	4	7	45
16-18	1	3	6	2	5	1	2	8	2	13	43
13-14	3	8	1	2	5	5		6	4	7	41
14-16	3	1	2	2	8		2	7	4	8	37

Table 3 (Continued)

DYS385	Lom	Ven	Lig	EmR	Tus	Mar	Umb	Lat	Pug	Sic	Total
11-13	1	5	2	3	15	1	1	1	3	4	36
14-15		5	1	4	2	3	3	9	3	3	33
13-17	1	1	1	3	7	2	3	11	1	2	32
14-14		6	4	1	3	1	1	4	4	6	30
12-16	3	1	6	3	4			4		5	26
13-13		6	3		7			3	2	4	25
14-17		2	4	1	5	2	2	5	1	3	25
11-16		4	1	7	3	3	2	3	1		24
12-15	1	4	1	2	6	1		4		4	23
15-15		2	2	2	2			8	2	5	23
17-18	1	1		1	3			4	3	9	22
15-18	1		1	2	2		2	3		8	19
15-16	1	2	2		3			4	1	5	18
16-17	1		1		2	1	1	5	1	6	18
13-18				1	2	1	2	2	1	8	17
11-11	2	1		1	4			7	1		16
12-13		3	1		4	2	1	4			15
16-16						3	1	3	1	7	15
14-18					2	1		2	3	4	12
15-17	1	1	1		1			2	1	5	12
16-19		1	1		1	1		7		1	12
10-14			1	2	3	2	1	2			11
12-12		1			4	1	1	3			10
12-17			1		5	1		3			10
17-17		1	2		2			2	2	1	10
13-19			1			1		6	1		9
10-13		5			3						8
11-17					3	2		2	1		8
10-17					2	1		1		1	5
12-18					1	1	1			2	5
11-12		1			1	1		1			4
19-21	1		2							1	4
13-20			1					1		1	3
14-19						1		1		1	3
15-19				1	1	1					3
17-19					1			1		1	3
17-20			1							2	3
18-19							1	1		1	3
09-12		1	1								2
09-13		2									2
10-11								1	1		2
10-15		1						1			2
10-16			2								2
16-21								1		1	2
09-15				1							1
09-17	1										1
10-10		1									1
10-12					1						1
10-18								1			1
11-19					1						1
12-19										1	1
12-20										1	1
15-20										1	1
16-20							1				1
18-20										1	1
19-19								1			1
Total	51	120	81	89	218	74	51	222	70	200	1176

Table 4
9-L haplotype frequency spectrum in 10 Italian regions and in the global sample

No. of individuals per haplotype	North				Center				South		Total
	Lom	Ven	Lig	EmR	Tus	Mar	Umb	Lat	Pug	Sic	
30											1
9											1
8											2
7											3
6					1						
5				1							3
4						1		1		1	13
3	1			1	3			2			21
2	0	5	2	7	10	4	1	10		11	54
1	48	110	77	67	183	62	49	192	70	174	862
No. of different haplotypes (<i>H</i>)	49	115	79	76	197	67	50	205	70	186	960
No. of individuals (<i>N</i>)	51	120	81	89	218	74	51	222	70	200	1176
<i>H/N</i>	0.96	0.96	0.98	0.85	0.90	0.91	0.98	0.92	1	0.93	0.82
Proportion of unique haplotypes	0.94	0.92	0.95	0.75	0.84	0.84	0.96	0.86	1	0.87	0.73

was 0.93, whereas in the total sample it was 0.82. The frequency of the most frequent haplotype in the entire sample ($n = 30$) was 0.026 (14-13-29-24-11-13-13-11-14). This haplotype was the most frequent in 8 out of 10 regions, and is also the more frequent in Iberian and Center-European populations (IFYUG on-line database). The proportion of the sample represented by unique haplotypes was 0.73 in the total sample and 0.93 in the regions.

6. Conclusions

A low degree of heterogeneity is present among Italian regions at the frequencies of Y-chromosome haplotypes; sporadic sampling problems and occasionally insufficient sub-sample size may in large part account for the observed diversity. More detailed population analyses may reveal micro-evolutionary trends throughout the Italian peninsula. A validated database of Y-chromosome haplotypes is now available for use in Court cases and in whatever forensic instance.

7. Access of data

www.gefi-forensicDNA.it

Acknowledgements

We acknowledge support from the following persons: Prof. M. Bacci (University di Perugia), Dr. G. Benucci

(University di Perugia), Prof. C. di Nunno (University of Bari), Dr. G. Forleo (University of Bari), Prof. G. Guanti (University of Bari), Dr. S. Prandoni (University of Milan), Dr. E. Rossi (University of Modena), Dr. D. Sangiorgi (University of Palermo), Dr. I. Sani (University of Florence), Prof. A. Tagliabracci (University of Ancona), Dr. C. Toni (University of Pisa), Prof. M.L.Giovannucci Uzielli (University of Florence).

References

- [1] M. Kayser, A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M.A. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, A. Piccinini, A. Perez-Lezaun, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, P. de Knijff, L. Roewer, Evaluation of Y-chromosomal STRs: a multicenter study, *Int. J. Legal Med.* 110 (1997) 125–133.
- [2] A. González-Neira, L. Gusmão, M. Brión, M.V. Lareu, A. Amorim, A. Carracedo, Distribution of Y-chromosome STR defined haplotypes in Iberia, *Forensic Sci. Int.* 110 (2000) 117–126.
- [3] A. Caglià, M. Dobosz, I. Boschi, E. d'Aloja, V.L. Pascali, Increased forensic efficiency of a STR-based Y-specific haplotype by addition of the highly polymorphic DYS385 locus, *Int. J. Legal Med.* 111 (1998) 142–146.
- [4] A. Helgason, S. Sigurðardóttir, J. Nicholson, B. Sykes, E.W. Hill, D.G. Bradley, V. Bosnes, J.R. Gulcher, R. Ward, K. Stefánsson, Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland, *Am. J. Hum. Genet.* 67 (2000) 697–717.