Int J Legal Med (2002) 116 : 139–147

ORIGINAL ARTICLE

Leonor Gusmão · Cíntia Alves · Sandra Beleza
António Amorim

Forensic evaluation and population data on the new Y-STRs DYS434, DYS437, DYS438, DYS439 and GATA A10

Received: 18 June 2001 / Accepted: 23 October 2001

Abstract The Y-specific STR loci DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS434, DYS437, DYS438, DYS439 and GATA A10 were studied in a northern Portuguese population. Haplotype and allele frequencies of these 14 Y-chromosome STRs were estimated. In a sample of 212 individuals it was possible to define 196 different haplotypes of which 182 were found only once, 12 were found in 2 samples and the 2 most frequent haplotypes were shared by only 3 individuals. The observed haplotype diversity value was 0.9992. The usefulness of the inclusion of each of these new markers for forensic purposes is discussed by comparing expected and observed increases in haplotype diversity. When combining the new markers (DYS434, DYS437, DYS438, DYS439 and GATA A10) with the classical set (DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393) a 0.68% increase in haplotype diversity was obtained and the number of different haplotypes rose from 157 to 196. When DYS434 was not considered the haplotype diversity was not affected.

Keywords North Portugal · Y-STRs · Haplotypes · Discrimination power · Haplotype diversity

Introduction

The use of Y-chromosome-specific STRs in population and forensic genetics has proved to be a very important tool in answering some specific questions (de Knijff et al. 1997; Jobling et al. 1997; Kayser et al. 1997a). In the forensic field these markers are especially helpful in deficient paternities and in rape cases (Prinz et al. 1997; Honda et al. 1999; Dekairelle and Hoste 2001; Corach et al. 2001).

With the increased use of these markers a high number of population databases have been produced (Roewer et al. 2001). The development of these databases is important not only for haplotype frequency estimation and direct application for match probability calculations in forensic studies but also to perform comparative population analysis (Roewer et al. 2000).

The Y-STR haplotype reference database is the most extensive survey on European populations available (internet site at http://ystr.charite.de). More recently, a second database has been generated including 28 regional U.S. population samples (Y-STR haplotype reference database for U.S. populations at http://ystr.org/usa).

These databases are very important for forensic users since it is well known that there are difficulties in the creation of large databases for the Y-linked markers, because the whole haplotype must be typed for each sample.

The choice of the STR core systems included in the Y-STR haplotype reference database of the international forensic Y-user group was made based on a restricted set of STRs (the ones that had been described at that time, well characterised and forensically validated; Kayser et al. 1997b). Later, more Y-STRs were described and used in the forensic field (White et al. 1999; Ayub et al. 2000; Hou et al. 2001; Gusmão et al. 2001a; González-Neira et al. 2001; Iida et al. 2001a, 2001b). According to Ayub et al. (2000) there are more than 150 Y-STR loci that may also be useful.

The addition of new Y-STRs to the previous set will allow an increase in the haplotype diversity value and consequently a higher power of discrimination. However, in routine forensic laboratory investigations this implies the need for re-constructing haplotypic databases (Pascali et al. 1998) as well as the development of new PCR multiplex strategies. For these reasons, it is important to quan-

L. Gusmão (✉) · C. Alves · S. Beleza · A. Amorim
IPATIMUP, Instituto de Patologia
e Imunologia Molecular da Universidade do Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
e-mail: lgusmao@ipatimup.pt,
Tel.: +351-22-5570700, Fax: +351-22-5570799

S. Beleza
Institute of Legal Medicine,
University of Santiago de Compostela, Spain

A. Amorim
Faculdade de Ciências, Universidade do Porto

Table 1 Allele frequencies at 14 Y-STRs in a northern Portuguese population (212 individuals)

DYS385			DYS19			DYS389I		
Haplotype	N	Frequency(± s.d.)	Allele	N	Frequency (± s.d.)	Allele	N	Frequency (± s.d.)
10–13	2	0.009 ± 0.007	13	21	0.099 ± 0.021	10	1	0.005 ± 0.005
10–14	2	0.009 ± 0.007	14	129	0.608 ± 0.034	11	2	0.009 ± 0.007
10–15	1	0.005 ± 0.005	15	50	0.236 ± 0.029	12	40	0.189 ± 0.027
11–11	1	0.005 ± 0.005	16	11	0.052 ± 0.015	13	126	0.594 ± 0.034
11–12	2	0.009 ± 0.007	17	1	0.005 ± 0.005	14	40	0.189 ± 0.027
11–13	7	0.033 ± 0.012	–	–	–	15	3	0.014 ± 0.008
11–14	81	0.382 ± 0.033	DYS389II			DYS390		
11–15	15	0.071 ± 0.018	Allele	N	Frequency (± s.d.)	Allele	N	Frequency (± s.d.)
11–16	3	0.014 ± 0.008	13	1	0.005 ± 0.005	21	5	0.024 ± 0.010
12–12	3	0.014 ± 0.008	15	14	0.066 ± 0.017	22	11	0.052 ± 0.015
12–13	3	0.014 ± 0.008	16	139	0.656 ± 0.033	23	45	0.212 ± 0.028
12–14	16	0.075 ± 0.018	17	45	0.212 ± 0.028	24	127	0.599 ± 0.034
12–16	3	0.014 ± 0.008	18	12	0.057 ± 0.016	25	23	0.108 ± 0.021
13–13	2	0.009 ± 0.007	19	1	0.005 ± 0.005	26	1	0.005 ± 0.005
13–14	11	0.052 ± 0.015	DYS391			DYS392		
13–15	7	0.033 ± 0.012	Allele	N	Frequency (± s.d.)	Allele	N	Frequency (± s.d.)
13–16	6	0.028 ± 0.011	08	1	0.005 ± 0.005	11	71	0.335 ± 0.032
13–17	4	0.019 ± 0.009	09	11	0.052 ± 0.015	12	10	0.047 ± 0.015
13–18	3	0.014 ± 0.008	10	103	0.486 ± 0.034	13	118	0.557 ± 0.034
13–19	1	0.005 ± 0.005	11	95	0.448 ± 0.034	14	11	0.052 ± 0.015
13–21	1	0.005 ± 0.005	12	2	0.009 ± 0.007	15	2	0.009 ± 0.007
14–14	9	0.042 ± 0.014	DYS393			DYS434		
14–15	7	0.033 ± 0.012	Allele	N	Freqency (± s.d.)	Allele	N	Freqency (± s.d.)
14–16	1	0.005 ± 0.005	12	36	0.170 ± 0.026	8	2	0.009 ± 0.007
14–17	1	0.005 ± 0.005	13	149	0.703 ± 0.031	9	207	0.976 ± 0.010
14–18	2	0.009 ± 0.007	14	23	0.108 ± 0.021	10	3	0.014 ± 0.008
14–19	1	0.005 ± 0.005	15	4	0.019 ± 0.009	–	–	–
15–15	5	0.024 ± 0.010	DYS437			DYS438		
15–16	2	0.009 ± 0.007	Allele	N	Frequency (± s.d.)	Allele	N	Frequency (± s.d.)
16–16	2	0.009 ± 0.007	14	67	0.316 ± 0.032	9	26	0.123 ± 0.023
16–18	1	0.005 ± 0.005	15	118	0.557 ± 0.034	10	52	0.245 ± 0.030
16–19	1	0.005 ± 0.005	16	27	0.127 ± 0.023	11	8	0.038 ± 0.013
17–17	2	0.009 ± 0.007	–	–	–	12	120	0.566 ± 0.034
17–19	1	0.005 ± 0.005	–	–	–	13	6	0.028 ± 0.011
17–20	1	0.005 ± 0.005	DYS439			GATA A10		
19–19	1	0.005 ± 0.005	Allele	N	Frequency (± s.d.)	Allele	N	Frequency (± s.d.)
12.2–14	1	0.005 ± 0.005	9	2	0.009 ± 0.007	13	9	0.042 ± 0.014
			10	23	0.108 ± 0.021	14	62	0.292 ± 0.031
			11	80	0.377 ± 0.033	15	109	0.514 ± 0.034
			12	83	0.392 ± 0.034	16	28	0.132 ± 0.023
			13	23	0.108 ± 0.021	17	3	0.014 ± 0.008
			14	1	0.005 ± 0.005	18	1	0.005 ± 0.005

titatively know the extra information obtained when further markers are added to a previously defined set.

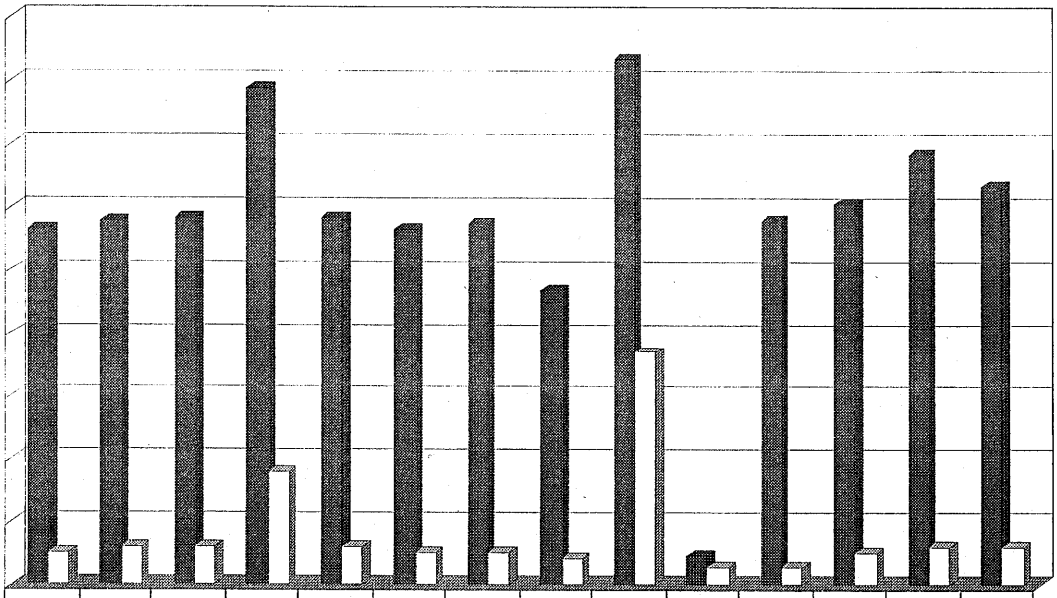
Since Y-STRs do not recombine, it is difficult to evaluate the potential information of a new STR just by the determination of the degree of polymorphism. Indeed, due to allelic association across loci, some STRs, although individually very polymorphic, may prove not to significantly increase the haplotype diversity previously obtained with core loci.

The aim of this work was to construct haplotypes including the nine STR loci, corresponding to the minimal haplotype of the Y-STR haplotype reference databases, plus 5 additional markers (DYS434, DYS437, DYS438, DYS439 and GATA A10), in order to evaluate the usefulness of their inclusion in forensic routine investigations.

Materials and methods

DNA samples

A sample of 212 unrelated healthy blood donors was selected from the north Portuguese population. Genomic DNA was extracted as described by Valverde et al. (1993).



	Y19	Y389I	Y389II	Y389	Y390	Y391	Y392	Y393	Y385	Y434	Y437	Y438	Y439	A10
Gene diversity	56.42	57.8	58.38	78.84	58.38	56.3	57.57	46.73	83.58	4.65	57.68	60.5	68.39	63.36
N. of alleles	5	6	6	18	6	5	5	4	37	3	3	5	6	6

Fig. 1 Gene diversities and allele numbers for the 14 Y-STR set

PCR amplification

Amplification was performed in four PCR reactions. DYS19, DYS389I and II, DYS390 and DYS393 were amplified in a pentaplex reaction as described by Gusmão et al. (1999). DYS391, DYS434, DYS437 and DYS439 were amplified in a PCR tetraplex reaction with 5 ng genomic DNA in a 12.5 µl reaction volume comprising 1.5 mM $MgCl_2$, 1 × buffer, 0.5 U Taq Gold polymerase (PE), 200 µM of each dNTP, 0.08 µM of DYS391 and DYS434 primers, 0.06 µM of DYS437 primers and 0.3 µM of DYS439 primers. A 95 °C pre-incubation step of 11 min was followed by 30 cycles at 94 °C denaturation for 30 s, annealing at 60 °C for 30 s, extension at 70 °C for 45 s and a 20 min final extension at 60 °C. Primer senquences were those described by Ayub et al. (2000) for DYS434, DYS437 and DYS439 and for DYS391 those described by Gusmão et al. (2000).

DYS385, DYS438 and GATA A10 were amplified in a triplex reaction in a 12.5 µl reaction volume comprising 1.5 mM $MgCl_2$, 1 × buffer, 0.5 U Taq Gold polymerase (Perkin Elmer), 200 µM of each dNTP, 0.24 µM of DYS385 primers, 0.3 µM of DYS438 primers and 0.12 µM of GATA A10 primers. A 95 °C pre-incubation step of 11 min was followed by 10 cycles at 94 °C denaturation for 30 s, annealing at 62 °C for 30 s, extension at 72 °C for 30 s and 20 cycles of 94 °C denaturation for 30 s, annealing at 60 °C for 30 s, extension at 72 °C for 30 s with 20 min final extension at 60 °C. DYS385 was amplified with the primers described by Schneider et al. (1998), DYS438 with those described by Ayub et al. (2000) and GATA A10 with those in White et al. (1999).

DYS392 was amplified in a singleplex reaction in a 12.5 µl reaction volume comprising 1.5 mM $MgCl_2$, 1 × buffer, 0.5 U Taq DNA polymerase recombinant (MBI Fermentas), 200 µM each dNTP and 0.2 µM of each primer. Amplifications were performed for 30 cycles of 30 s at 94 °C, 30 s at 58 °C and 1 min at 72 °C. The primers used were those in the Human Genome Data Base (Locus GDB-ID G00–456–509).

Detection system

For genetic typing, ABI310 and ABI377 automatic sequencers (Perkin-Elmer) along with the Genescan 2.1 analysis software were used.

Allele designations were based on comparison with the allelic ladders obtained by the mixture of previously sequenced samples for the most common alleles. The allele nomenclature was as proposed by Kayser et al. (1997 a) with the exception of the DYS389I where three monomorphic repeats were added (according to the nomenclature in the Y-STR haplotype reference databases) and DYS389II where alleles were named by subtracting the DYS389I stretch from the total number of repeats. DYS434, DYS437, DYS438 and DYS439 alleles were named according to Gusmão et al. (2001 a) and GATA A10 as $(TCCA)_2(TATC)_n$ according to Gusmão et al. (2001 b).

Statistical analysis

Allele and haplotype frequencies were estimated by gene or haplotype counting and observed gene and haplotype diversities were estimated according to Nei (1987). We defined expected haplotype diversity as the theoretical value calculated assuming no association between each of the alleles at a specific locus and the previously defined core loci haplotypes.

Population differentiation was tested by a Markov test using the Arlequin software ver. 2000 (Schneider et al. 2000).

Results and discussion

Single locus analysis

Allele frequencies at each STR locus are shown in Table 1. A smaller sample from the same northern Portuguese population was previously studied for DYS19, DYS389I and II, DYS390, DYS391, DYS392, DYS393, DYS434, DYS437, DYS438 and DYS439 loci (González-Neira

Table 2 Y chromosome haplotypes in 212 individuals from a northern Portuguese population (the nomenclature for DYS389II in parenthesis is in accordance with the Y-STR haplotype reference databases)

Haplotype	N	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 434	DYS 437	DYS 438	DYS 439	GATA A10
H1	1	13	12	17 (29)	24	10	11	13	16/16	9	14	10	13	14
H2	2	13	12	17 (29)	24	10	11	13	17/17	9	14	10	10	15
H3	1	13	12	17 (29)	24	10	11	13	19/19	9	14	10	10	15
H4	1	13	13	16 (29)	24	9	11	13	13/13	9	14	10	11	14
H5	2	13	13	16 (29)	24	9	11	13	13/14	9	14	10	10	14
H6	1	13	13	16 (29)	24	9	11	13	13/14	9	14	10	11	14
H7	1	13	13	16 (29)	24	11	13	12	11/15	9	15	12	12	16
H8	1	13	13	16 (29)	25	9	11	13	13/14	9	14	10	10	15
H9	1	13	13	17 (30)	24	10	11	12	16/19	9	14	10	12	15
H10	1	13	13	17 (30)	25	10	11	13	14/18	10	14	10	13	14
H11	1	13	13	18 (31)	23	10	11	13	16/16	9	14	10	12	15
H12	1	13	13	18 (31)	24	10	11	13	16/18	9	14	10	11	13
H13	2	13	14	16 (30)	24	9	11	13	13/14	9	14	10	10	14
H14	1	13	14	16 (30)	24	9	11	13	14/14	9	14	10	10	14
H15	1	13	14	16 (30)	24	9	11	13	14/14	9	14	10	11	14
H16	1	13	14	16 (30)	24	11	11	12	13/18	9	16	9	13	14
H17	1	13	15	16 (31)	24	9	11	13	13/14	9	14	10	10	14
H18	1	13	15	16 (31)	25	11	13	13	12/14	9	15	12	12	15
H19	1	14	10	16 (26)	24	11	14	13	12/14	9	15	12	12	16
H20	1	14	11	15 (26)	24	11	13	13	11/14	9	15	12	13	14
H21	1	14	12	16 (28)	22	10	11	13	13/14	9	16	10	11	15
H22	1	14	12	16 (28)	22	10	11	13	13/15	9	16	10	11	15
H23	1	14	12	16 (28)	22	10	11	13	14/14	9	16	10	11	16
H24	1	14	12	16 (28)	23	10	11	13	13/15	9	16	10	11	15
H25	1	14	12	16 (28)	23	10	13	13	12/14	9	15	12	12	15
H26	1	14	12	16 (28)	24	10	13	13	10/14	9	14	12	13	15
H27	1	14	12	16 (28)	24	10	13	13	12/14	9	15	12	12	15
H28	1	14	12	16 (28)	24	11	11	13	13/16	9	16	10	11	15
H29	1	14	12	16 (28)	24	11	13	13	11/14	9	15	12	13	16
H30	1	14	12	16 (28)	24	11	13	13	11/15	9	15	12	12	15
H31	1	14	12	16 (28)	25	10	12	13	11/14	9	15	12	11	15
H32	1	14	12	16 (28)	25	11	13	13	11/14	9	15	12	12	14
H33	1	14	12	16 (28)	25	11	13	13	11/14	9	15	12	13	15
H34	1	14	12	18 (30)	24	11	13	13	11/14	9	14	12	11	13
H35	1	14	13	13 (26)	23	10	13	13	13/16	9	14	9	11	14
H36	1	14	13	15 (28)	23	10	13	13	10/13	9	15	12	11	14
H37	1	14	13	15 (28)	23	10	13	13	11/14	9	15	12	11	15
H38	1	14	13	15 (28)	24	10	13	13	11/14	9	15	12	14	15
H39	1	14	13	15 (28)	24	10	13	13	12/14	9	15	13	11	15
H40	1	14	13	15 (28)	24	11	13	13	11/13	9	15	12	12	15
H41	1	14	13	15 (28)	25	10	13	13	12/14	9	15	12	11	15
H42	1	14	13	15 (28)	25	11	13	13	11/14	9	15	12	11	15
H43	1	14	13	15 (28)	26	11	13	12	11/14	9	15	12	12	14
H44	1	14	13	16 (29)	23	10	11	12	14/16	9	14	9	10	15
H45	1	14	13	16 (29)	23	10	13	13	11/15	9	15	12	11	15
H46	2	14	13	16 (29)	23	10	13	13	13/15	9	14	9	11	14
H47	1	14	13	16 (29)	23	11	13	13	11/13	9	15	12	13	15
H48	1	14	13	16 (29)	23	11	13	13	11/14	8	15	12	12	14
H49	1	14	13	16 (29)	23	11	13	13	11/14	9	14	12	11	16
H50	1	14	13	16 (29)	23	11	13	13	11/14	9	14	12	12	15
H51	1	14	13	16 (29)	23	11	13	13	11/14	9	15	12	11	15
H52	1	14	13	16 (29)	23	11	13	13	11/14	9	15	12	12	14
H53	1	14	13	16 (29)	23	11	13	13	11/14	9	15	12	12	15
H54	1	14	13	16 (29)	23	11	13	13	11/14	9	15	13	12	14
H55	1	14	13	16 (29)	23	11	13	13	11/14	10	15	12	11	15
H56	1	14	13	16 (29)	23	11	13	13	12/14	9	16	12	11	15

Table 2 (continued)

Haplo-type	N	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 434	DYS 437	DYS 438	DYS 439	GATA A10
H57	1	14	13	16 (29)	23	11	14	13	11/14	9	14	12	10	14
H58	1	14	13	16 (29)	24	10	13	12	11/14	9	15	12	11	16
H59	1	14	13	16 (29)	24	10	13	13	10/14	9	15	12	12	15
H60	2	14	13	16 (29)	24	10	13	13	11/14	9	14	12	11	15
H61	1	14	13	16 (29)	24	10	13	13	11/14	9	14	12	12	16
H62	1	14	13	16 (29)	24	10	13	13	11/14	9	15	12	11	15
H63	1	14	13	16 (29)	24	10	13	13	11/14	9	15	12	12	14
H64	1	14	13	16 (29)	24	10	13	13	11/14	9	15	12	12	15
H65	1	14	13	16 (29)	24	10	13	13	11/15	9	15	12	12	15
H66	2	14	13	16 (29)	24	10	13	13	12/14	9	15	12	12	14
H67	1	14	13	16 (29)	24	10	13	13	12/14	10	14	12	12	14
H68	1	14	13	16 (29)	24	10	13	14	11/14	9	16	12	12	15
H69	1	14	13	16 (29)	24	10	13	14	11/14	9	16	12	13	15
H70	1	14	13	16 (29)	24	10	14	13	11/14	9	15	12	11	15
H71	1	14	13	16 (29)	24	11	12	13	11/14	9	15	12	12	16
H72	2	14	13	16 (29)	24	11	13	12	11/14	9	15	12	12	15
H73	1	14	13	16 (29)	24	11	13	12	11/15	9	15	12	10	14
H74	1	14	13	16 (29)	24	11	13	12	11/15	9	15	12	12	16
H75	1	14	13	16 (29)	24	11	13	13	11/13	9	14	12	11	15
H76	1	14	13	16 (29)	24	11	13	13	11/14	9	14	12	12	14
H77	2	14	13	16 (29)	24	11	13	13	11/14	9	14	12	12	15
H78	1	14	13	16 (29)	24	11	13	13	11/14	9	15	10	12	15
H79	1	14	13	16 (29)	24	11	13	13	11/14	9	15	12	9	15
H80	2	14	13	16 (29)	24	11	13	13	11/14	9	15	12	11	15
H81	2	14	13	16 (29)	24	11	13	13	11/14	9	15	12	11	16
H82	1	14	13	16 (29)	24	11	13	13	11/14	9	15	12	12	14
H83	1	14	13	16 (29)	24	11	13	13	11/14	9	15	12	12	16
H84	1	14	13	16 (29)	24	11	13	13	11/14	9	15	12	13	14
H85	1	14	13	16 (29)	24	11	13	13	11/14	9	15	12	13	15
H86	1	14	13	16 (29)	24	11	13	13	12/12	9	14	12	11	15
H87	1	14	13	16 (29)	24	11	13	13	12/14	9	14	11	11	14
H88	1	14	13	16 (29)	24	11	13	13	12/14	9	15	12	12	14
H89	1	14	13	16 (29)	24	11	13	13	12/14	9	15	13	13	14
H90	1	14	13	16 (29)	24	11	13	14	11/13	9	15	12	12	14
H91	1	14	13	16 (29)	24	11	13	14	11/14	9	15	12	11	15
H92	1	14	13	16 (29)	24	11	14	13	11/14	9	14	12	11	14
H93	1	14	13	16 (29)	25	10	11	12	12/13	9	15	9	11	15
H94	1	14	13	16 (29)	25	10	13	14	11/14	9	15	12	11	15
H95	1	14	13	16 (29)	25	10	14	13	11/14	9	15	12	13	14
H96	1	14	13	16 (29)	25	11	13	13	11/13	9	15	12	12	15
H97	1	14	13	16 (29)	25	11	13	13	11/14	9	15	12	12	14
H98	1	14	13	16 (29)	25	12	13	13	11/12	9	14	12	12	15
H99	1	14	13	17 (30)	23	8	11	13	13/14	9	14	10	10	14
H100	1	14	13	17 (30)	23	10	11	12	13/18	9	14	10	12	15
H101	1	14	13	17 (30)	23	10	13	13	11/15	9	15	12	11	15
H102	1	14	13	17 (30)	24	10	11	12	14/15	9	15	9	12	15
H103	1	14	13	17 (30)	24	10	13	13	10/13	9	15	12	11	16
H104	1	14	13	17 (30)	24	10	13	13	11/14	9	15	12	12	15
H105	1	14	13	17 (30)	24	11	13	13	11/12	9	15	12	11	15
H106	1	14	13	17 (30)	24	11	13	13	11/15	9	14	12	12	15
H107	1	14	13	17 (30)	24	11	14	13	12/14	9	15	12	12	15
H108	1	14	13	17 (30)	24	12	11	12	13/19	9	14	10	10	16
H109	1	14	13	17 (30)	25	11	14	13	11/14	9	14	12	12	15
H110	1	14	13	18 (31)	24	10	11	13	17/20	9	14	10	12	15
H111	1	14	13	18 (31)	24	10	13	13	11/14	9	15	12	12	15
H112	1	14	13	18 (31)	25	10	11	12	14/14	9	14	10	11	15
H113	1	14	14	15 (29)	24	10	14	12	12.2/14	9	15	12	13	15

Table 2 (continued)

Haplo-type	N	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 434	DYS 437	DYS 438	DYS 439	GATA A10
H114	1	14	14	16 (30)	23	10	11	12	13/13	9	14	10	11	14
H115	1	14	14	16 (30)	23	10	15	12	17/19	9	16	10	11	14
H116	1	14	14	16 (30)	24	10	13	13	11/15	9	14	12	11	14
H117	1	14	14	16 (30)	24	11	13	13	11/14	9	14	12	12	17
H118	1	14	14	16 (30)	24	11	13	13	11/14	9	15	11	13	13
H119	1	14	14	16 (30)	24	11	13	13	11/14	9	15	12	11	14
H120	1	14	14	16 (30)	24	11	13	13	11/14	9	15	12	11	15
H121	2	14	14	16 (30)	24	11	13	13	11/14	9	15	12	12	15
H122	3	14	14	16 (30)	24	11	13	13	11/15	9	15	12	12	15
H123	1	14	14	16 (30)	24	11	13	13	12/14	9	15	12	12	15
H124	1	14	14	16 (30)	24	11	13	14	11/14	9	15	12	12	15
H125	1	14	14	16 (30)	25	10	11	12	14/18	9	14	10	11	14
H126	1	14	14	16 (30)	25	10	13	13	11/14	8	15	12	12	15
H127	1	14	14	16 (30)	25	10	13	13	11/14	9	15	13	13	16
H128	1	14	14	16 (30)	25	10	14	13	11/14	9	15	12	13	14
H129	1	14	14	17 (31)	23	10	11	12	14/15	9	14	10	11	16
H130	1	14	14	17 (31)	23	10	11	13	14/15	9	14	9	11	15
H131	1	14	14	17 (31)	24	10	13	13	15/16	9	14	9	11	13
H132	3	14	14	17 (31)	24	11	13	13	11/14	9	15	12	12	15
H133	1	14	14	18 (32)	23	10	11	12	12/13	9	14	12	12	15
H134	1	14	14	18 (32)	23	10	11	13	14/15	9	14	9	12	14
H135	1	14	15	15 (30)	24	11	13	13	11/14	9	15	12	12	14
H136	1	15	11	15 (26)	24	10	13	14	10/15	9	15	12	12	15
H137	1	15	12	15 (27)	24	10	14	13	13/16	9	14	9	11	15
H138	1	15	12	16 (28)	21	10	11	14	13/15	9	16	11	11	13
H139	1	15	12	16 (28)	21	10	11	14	15/15	9	16	10	12	15
H140	1	15	12	16 (28)	21	10	11	15	13/15	9	15	10	11	15
H141	1	15	12	16 (28)	23	10	11	14	14/14	9	16	10	12	14
H142	1	15	12	16 (28)	24	10	11	12	13/17	9	16	9	12	14
H143	1	15	12	16 (28)	24	10	11	13	13/17	9	16	9	11	14
H144	1	15	12	16 (28)	24	10	11	13	14/17	9	16	9	11	14
H145	1	15	12	17 (29)	21	10	11	13	14/15	9	16	10	11	15
H146	1	15	12	17 (29)	21	10	11	15	13/17	9	16	10	12	13
H147	2	15	12	17 (29)	22	10	11	14	15/15	9	16	10	12	15
H148	1	15	12	17 (29)	23	10	11	12	13/16	9	15	9	10	15
H149	1	15	12	17 (29)	23	10	12	14	13/15	9	15	10	9	14
H150	1	15	12	17 (29)	24	11	13	13	12/14	9	15	12	10	15
H151	1	15	12	17 (29)	25	11	11	13	11/14	9	15	12	13	15
H152	1	15	12	18 (30)	23	10	12	14	14/14	9	16	9	12	14
H153	1	15	12	18 (30)	24	10	12	14	13/16	9	16	10	11	15
H154	1	15	13	16 (29)	22	11	11	12	12/16	9	15	9	11	15
H155	1	15	13	16 (29)	22	11	13	13	11/14	9	15	12	13	15
H156	1	15	13	16 (29)	23	9	11	13	13/16	9	14	9	11	14
H157	1	15	13	16 (29)	23	10	11	12	13/21	9	15	9	11	16
H158	1	15	13	16 (29)	23	10	11	12	14/14	9	15	9	11	14
H159	1	15	13	16 (29)	23	11	13	12	11/14	9	15	12	12	15
H160	1	15	13	16 (29)	24	10	11	12	13/18	9	15	9	12	16
H161	1	15	13	16 (29)	24	10	12	14	11/14	9	15	12	11	15
H162	1	15	13	16 (29)	24	10	13	13	11/15	9	15	12	12	15
H163	1	15	13	16 (29)	24	10	13	13	11/15	9	15	12	12	16
H164	1	15	13	16 (29)	24	10	13	13	11/16	9	15	12	12	15
H165	1	15	13	16 (29)	24	11	11	12	14/19	9	15	9	12	15
H166	1	15	13	16 (29)	24	11	13	12	11/16	9	14	12	12	15
H167	1	15	13	16 (29)	24	11	13	13	11/13	9	14	12	11	15
H168	1	15	13	16 (29)	24	11	13	13	11/14	9	15	12	12	13
H169	1	15	13	16 (29)	24	11	13	13	11/14	9	15	12	12	14
H170	1	15	13	16 (29)	24	11	13	13	11/14	9	15	12	13	16

Table 2 (continued)

Haplo-type	N	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 434	DYS 437	DYS 438	DYS 439	GATA A10
H171	1	15	13	16 (29)	25	10	13	12	11/14	9	15	13	11	14
H172	1	15	13	17 (30)	22	11	11	13	14/14	9	16	10	12	14
H173	1	15	13	17 (30)	23	10	12	15	15/16	9	14	10	11	16
H174	1	15	13	17 (30)	23	11	11	12	12/16	9	15	9	13	13
H175	1	15	13	17 (30)	23	11	11	12	12/16	9	15	9	13	16
H176	1	15	13	17 (30)	24	10	11	12	13/17	9	14	10	11	14
H177	1	15	13	17 (30)	24	10	12	15	15/15	9	15	10	11	15
H178	1	15	13	17 (30)	24	11	13	13	11/13	9	14	12	12	14
H179	1	15	13	17 (30)	24	11	13	14	11/14	9	15	13	11	16
H180	1	15	14	16 (30)	24	10	14	12	11/15	9	15	12	11	15
H181	1	15	14	16 (30)	24	11	13	13	11/14	9	15	12	11	16
H182	1	15	14	16 (30)	24	11	13	13	11/14	9	15	12	12	15
H183	1	15	14	17 (31)	23	11	11	12	12/13	9	16	9	11	15
H184	1	15	14	17 (31)	24	10	15	13	11/16	9	15	12	13	15
H185	1	16	12	16 (28)	22	10	11	14	13/14	9	16	10	11	14
H186	1	16	12	16 (28)	22	10	11	14	14/15	9	16	11	11	13
H187	1	16	12	17 (29)	22	11	11	13	13/14	9	16	10	11	14
H188	1	16	12	17 (29)	24	10	12	14	14/14	9	15	10	10	16
H189	1	16	13	16 (29)	24	11	12	14	15/15	9	15	10	11	15
H190	1	16	13	18 (31)	24	10	11	14	11/14	9	14	11	10	16
H191	1	16	13	18 (31)	25	11	11	13	11/14	9	14	11	10	16
H192	1	16	13	19 (32)	24	10	11	13	11/11	9	14	11	10	17
H193	1	16	14	16 (30)	24	10	11	14	11/14	9	14	11	10	16
H194	1	16	14	17 (31)	23	10	13	13	14/15	9	14	9	10	15
H195	1	16	14	17 (31)	23	11	11	13	12/12	9	15	10	11	17
H196	1	17	13	15 (28)	23	10	11	13	12/12	9	15	10	11	16

et al. 2000; Gusmão et al. 2001a). For most loci, the increase of sample size led to an increase of gene diversity (Fig. 1).

For the five new markers (DYS434, DYS437, DYS438, DYS439 and GATA A10), gene diversity values are in the same range as those found for the classical set or even higher, with the exception of DYS434 (in accordance with Gusmão et al. 2001a).

Comparing gene diversity and allele number for all 14 Y-STRs, we can see that the average increase in gene diversity correlates well with the increase in allele number (Fig. 1). However, this is not the case for DYS434 and DYS437 which share the same number of aleles (3) but have very different gene diversity values. Moreover, DYS437 has a very high gene diversity value in the range of the loci with five alleles.

Haplotype analysis

A list of the haplotypes is given in Table 2. In a sample of 212 individuals it was possible to define 196 different haplotypes of which 182 were found only once, 12 were shared by 2 individuals and the 2 most frequent haplotypes were shared by only 3 individuals.

The observed haplotype diversity value was 0.9992 (Table 3). In the same sample, when combining the new markers with the classical set (DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393) a 0.68% increase in haplotype diversity was obtained and the number of different haplotypes rose from 157 to 196.

Forensic assessment

Due to allelic association across loci some STRs, although individually very polymorphic, may prove to insignificantly increase the haplotype diversity previously obtained with core loci. In other words, due to the non-recombining nature of the chromosome region under analysis it can happen that for an individually very informative marker, its alleles are not able to discriminate more haplotypes than those previously defined by the “core” loci, and therefore this extra information becomes redundant.

The usefulness of the addition of each of the five new markers to the previously defined nine loci set included in the Y-STR haplotype reference databases was evaluated.

The haplotype diversity values were calculated adding each of the five additional markers to the nine Y-STR core loci (Table 3). All the markers contributed to an increase in the number of different haplotypes.

DYS439 and GATA A10 allowed the discrimination of a higher number of different haplotypes, with 175 haplotypes being discriminated, compared with the 157 obtained with the classical set. However, DYS439 led to a higher haplotype diversity than GATA A10. The same

Table 3 Haplotype number and diversity values in a sample of 212 individuals from a northern Portuguese population by combining different Y-STR sets

Haplotype composition	Number of different haplotypes	Haplotype diversity (%)
9 Y-STR set[a]	157 (74.06%)	99.25
9 Y-STR set plus DYS434	162	99.32
9 Y-STR set plus DYS437	164	99.52
9 Y-STR set plus DYS438	164	99.38
9 Y-STR set plus DYS439	175	99.72
9 Y-STR set plus GATA A10	175	99.70
14 Y-STR set[b]	196 (92.45%)	99.92
14 Y-STR set without DYS434	196	99.92
14 Y-STR set without DYS437	193	99.90
14 Y-STR set without DYS438	195	99.92
14 Y-STR set without DYS439	186	99.85
14 Y-STR set without GATA A10	186	99.86

[a] 9 loci set included in the Y-STR haplotype reference databases.
[b] 14 Y-STR core studied in this work.

happened with the addition of DYS437 or DYS438, where different haplotype diversity values were obtained for the same number of haplotypes.

The theoretical haplotype diversity expected was calculated when each of the five markers was added and compared with the observed value (Fig. 2). Since we defined expected haplotype diversity as the theoretical value calculated as if the locus under analysis was totally unlinked to the core haplotype, this analysis allows the comparison between the specific locus individual informativeness and its effective contribution to the observed haplotype diversity. For most loci the observed increase was lower than expected and was much lower for DYS438. On the contrary DYS434, although contributing less, showed an observed increase much higher than expected. The results of this analysis confirmed that it is impossible to evaluate the usefulness of a Y-STR system outside an empirical haplotype approach.

The informative contribution of DYS434, DYS437, DYS438, DYS439 and GATA A10 to an extended Y-STR database was evaluated by calculating the haplotype diversity values excluding each of the five additional markers from the whole 14 Y-STR set (Table 3).

When DYS434 and DYS438 were not considered, the haplotype diversity was not affected. A decrease in the haplotype diversity was obtained when excluding any of the other three markers (DYS437, DYS439 and GATA A10).

Comparing these loci, it is possible to conclude that the increase in haplotype diversity is not directly correlated to gene diversity. Indeed, when DYS438 was not considered the haplotype diversity was the same as for the whole 14 STR core and the number of different haplotypes was only slightly affected with the loss of just one haplotype although this system has a high gene diversity (0.6050). With the inclusion of DYS437, the haplotype diversity was barely affected (increasing from 0.9990 to 0.9992), with three more haplotypes being distinguished.

The markers that proved to be most useful for inclusion in the forensic routine were DYS439 and GATA A10. When each of these markers were added to the database, there was an increase in the haplotype diversity to 0.9992 from 0.9985 and 0.9986, respectively. Although these two markers have different gene diversity values (higher for DYS439), both contributed to the same increase in the number of different haplotypes (from 186 to 196).

Another interesting result was obtained when comparing DYS385 and DYS389, the most informative markers in the studied population. In the context of the markers considered in this work, although with different number of allelic classes (18 for DYS389I and II and 37 for DYS385) and gene diversities (0.7884 for DYS389I and II and 0.8358 for DYS385), they have the same impact on the haplotype diversity. When each of these STRs is not considered the number of different haplotypes decreases from 196 to 183 in both cases (data not shown).

In conclusion, we can state that in the context of this population DYS437, DYS438, DYS439 and GATA A10 contributed to a higher power of discrimination when added to the previous set of Y-STRs (those included in the Y-STR haplotype reference database). The inclusion of DYS434 in the forensic routine will not raise the power of discrimination and, moreover, with the marginal disadvantage of an increase of the mutation rate expected for the whole haplotype (Kayser and Sajantila 2001).

It must be stressed that the conclusions reported here cannot be extrapolated to other (in particular, genetically distant) populations. In fact, not only distinct gene diver-

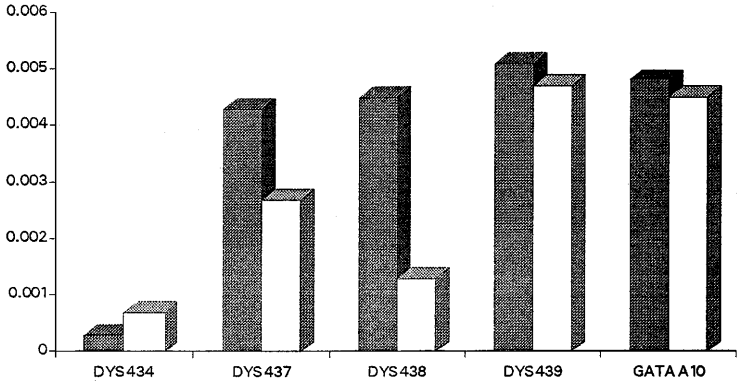


Fig. 2 Observed and expected increase in haplotype diversity by the addition of new Y-STRs to the Y-reference database core loci haplotypes in North Portugal (*solid columns* expected increase, *empty columns* observed increase)

sities per locus can occur, but more importantly, different associations between loci may also occur .

Population comparison

Recently Grignani et al. (2000) reported the results of the haplotype diversity in a sample from northwest Italy, by the addition of the three STRs, DYS437, DYS438 and DYS439 to the previously established nine Y-STR set. In order to compare our data for northern Portugal with those published by Grignani et al. (2000), we reanalysed our data excluding DYS434 and GATA A10. We found the same haplotype diversity values in both samples (0.9984 ± 0.0007 in north Portugal and 0.9985 ± 0.0014 to northwestern Italy). By combining both samples (340 individuals) 304 haplotypes are distinguished and only 6 are shared by both populations. The AMOVA results show that the percentage of variation is mainly within populations (99.35%) in agreement with previous results in European populations (Roewer et al. 2000).

Acknowledgements This work was partially supported by Fundação para a Ciência e a Tecnologia (through grant SFRH/BD/860/2000 and POCTI, Programa Operacional Ciência, Tecnologia e Inovação)

References

Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res 2:e8

Corach D, Filgueira Risso L, Marino M, Penacino G, Sala A (2001) Routine Y-STR typing in forensic casework. Forensic Sci Int 118:133–138

Dekairelle AF, Hoste B (2001) Application of a Y-STR-pentaplex PCR (DYS19, DYS389I and II, DYS390 and DYS393) to sexual assault cases. Forensic Sci Int 118:122–125

González-Neira A, Gusmão L, Brión M, Lareu MV, Amorim A, Carracedo A (2000) Distribution of Y-chromosome STR defined haplotypes in Iberia. Forensic Sci Int 110:117–126

González-Neira A, Elmoznino M, Lareu MV, Sánchez-Diz P, Gusmão L, Prinz M, Carracedo A (2001) Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies. Forensic Sci Int (in press)

Grignani P, Peloso G, Fattorini P, Previderè C (2000) Highly informative Y-chromosomal haplotypes by the addition of three new STRs DYS437, DYS438 and DYS439. Int J Legal Med 114:125–129

Gusmão L, González-Neira A, Pestoni C, Brión M, Lareu MV, Carracedo A (1999) Robustness of the Y STRs DYS19, DYS389I and II, DYS390 and DYS393:optimization of a PCR pentaplex. Forensic Sci Int 106:163–172

Gusmão L, González-Neira A, Sánchez-Diz P, Lareu MV, Amorim A, Carracedo A (2000) Alternative primers for DYS391 typing: advantages of their application to forensic genetics. Forensic Sci Int 112:49–57

Gusmão L, Alves C, Amorim A (2001a) Molecular characterization of four Y-specific microsatellites (DYS434, DYS437, DYS438, DYS439) for population and forensic studies. Ann Hum Genet 65:285–291

Gusmão L, Gonzaléz-Neira A, Alves C, Amorim A, Carracedo A (2001b) Y chromosome short tandem repeat polymorphisms: a comparison between humans and chimpanzees. In: Progress in Forensic Genetics 9. Elsevier, Amsterdam (in press)

Honda K, Roewer L, Knijff P de (1999) Male DNA typing from 25-year-old vaginal swabs using Y chromosomal STR polymorphisms in a retrial request case. J Forensic Sci 44:868–872

Hou YP, Zhang J, Li YB, Wu J, Zhang SZ, Prinz M (2001) Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population. Forensic Sci Int 118:153–159

Iida R, Tsubota E, Matsuki T (2001a) Identification and characterization of two novel human polymorphic STRs on the Y chromosome. Int J Legal Med 115:54–56

Iida R, Tsubota E, Sawazaki K, Masuyama M, Matsuki T, Yasuda T, Kishi K (2001b) Characterization and haplotype analysis of the polymorphic Y-STRs, DYS443, DYS444 and DYS445 in a Japanese population. Int J Legal Med (in press)

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110: 118–124

Kayser M, Sajantila A (2001) Mutation at Y-STR loci: implications for paternity testing and forensic analysis. Forensic Sci Int 118:116–121

Kayser M, Knijff P de, Dieltjes P, et al, (1997a) Applications of microsatellite-based Y chromosome haplotyping. Electrophoresis 18:1602–1607

Kayser M, Cagliá A, Corach D, et al, (1997b) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110: 125–133; appendix 141–149

Knijff P de, Kayser M, Caglià A, et al, (1997) Chromosome Y microsatellites: population genetics and evolutionary aspects. Int J Legal Med 110:134–149; appendix 141–149

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Pascali VL, Dobosz M, Brinkmann B (1998) Coordinating Y-chromosomal STR research for the courts. Int J Legal Med 112:1

Prinz M, Boll K, Baum H, Shaler B (1997) Multiplexing of Y chromosome specific STRs and performance for mixed samples. Forensic Sci Int 85:209–218

Roewer L, Kayser M, Knijff P de, et al, (2000) A new method for evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. Forensic Sci Int 114:31–43

Roewer L, Krawczak M, Willuweit S, et al, (2001) Online reference database of Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106–113

Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998) Tandem repeat structure of the duplicated Y chromosomal STR locus DYS385 and frequency studies in the German and Asian populations. Forensic Sci Int 97:61–70

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000. A software for population genetics data analysis. University of Geneva

Valverde E, Cabrero C, Cao R (1993) Population genetics of three VNTR polymorphisms in two different Spanish populations. Int J Legal Med 151:251–256

White PS, Tatum OL, Deaven LL, Longmire JL (1999) New, male-specific microsatellite markers from the human Y chromosome. Genomics 57:433–437