

High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions

Elena Bosch^a, Andrew C. Lee^a, Francesc Calafell^b, Eduardo Arroyo^c, Peter Henneman^d, Peter de Knijff^d, Mark A. Jobling^{a,*}

^aDepartment of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

^bUnitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain

^cFacultad de Medicina, Universidad Complutense de Madrid, Madrid, Spain

^dForensic Laboratory for DNA-Research, MGC-Department of Human and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands

Received 30 July 2001; accepted 26 October 2001

Abstract

Nineteen Y-specific short tandem repeat (STR) loci have been amplified in 768 samples from the Iberian Peninsula in order to evaluate their usefulness in forensic casework. Two previously published multiplex reactions by Thomas et al. [Hum. Genet. 6 (1999) 577] (MS1, modified here: DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393) and by Ayub et al. [Nucl. Acids Res. 28 (2000) e8] (CTS: DYS434, DYS435, DYS436, DYS437, DYS438 and DYS439) plus a novel one reported here (EBF: DYS385, DYS389, DYS460, DYS461, DYS462 and amelogenin) have been used. DYS385, DYS439 and DYS391 were the most informative loci with allele diversities of 0.7997, 0.6683 and 0.5940, respectively. A total of 635 different haplotypes were observed, of which 573 (90.24%) were found in single individuals. The overall haplotype diversity was 0.9988 and that obtained by each multiplex system was 0.9812 for EBF, 0.9292 for MS1 and 0.9089 for CTS. © 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Y chromosome; STR polymorphisms; Microsatellites; Haplotypes; Iberian Peninsula; Multiplex

1. Introduction

The use of Y chromosome markers is becoming common in forensic genetics because of their male-specificity [1]. They are the markers of choice in sex crimes in which one or more males are involved, since they obviate the need for differential separation of sperm DNA from a mixed sample. The most informative markers for individual identification on the Y chromosome are the MSY1 minisatellite [2] and a number of short tandem repeats (STRs) [3,4]. It should be noted that, given the haploid nature of the Y chromosome, each individual Y-STR (of which each male will normally carry one allele) would be less informative than an autosomal counterpart having the same allele diversity, but displaying two alleles per individual. Furthermore, because of the lack of recombination, haplotype diversity is lower than that of a

segment of recombining autosome of a similar size, and with similar STR density. Thus, a larger number of Y chromosome STRs is needed to achieve a power of discrimination (PD) similar to that of autosomal STR sets. The availability of a number of Y-STRs, preferably amplified and typed in a small number of multiplex reactions, will thus be a valuable toolkit in forensic genetics. The establishment of multiplex systems has the further advantage of consuming less template DNA and reagents, reducing labour, and more importantly, reducing the chances of contamination and/or error.

Recently, a number of new Y-specific STRs have been developed and studied in different populations [5–9]. Given that the number of loci available is growing, efficient new methods are required for their amplification and typing. Moreover, the number and geographic distribution of the population groups tested so far for these new markers are limited. Further characterisation of the allele frequency distribution and pattern of variation of the new markers is necessary in order to evaluate their usefulness in forensic applications.

* Corresponding author. Tel.: +44-116-252-3377; fax: +44-116-252-3378. E-mail address: maj4@leicester.ac.uk (M.A. Jobling).

Here, we present a new multiplex reaction and population data on 19 Y-specific STRs (DYS19, DYS385, DYS389I/AB, DYS388, DYS390, DYS391, DYS392, DYS393, DYS434, DYS435, DYS436, DYS437, DYS438, DYS439, DYS460, DYS461 and DYS462) in the Iberian Peninsula. We (1) evaluate the number of haplotypes produced as well as the PD obtained by progressively combining the different loci, (2) evaluate the additional haplotype diversity achieved by adding the extra loci to the minimal 9-locus 'core' haplotype used for the Y-STR Haplotype Reference Database (YHRD: <http://www.ystr.org>; [10]) ranked by their diversity values, and (3) compare the haplotype diversity obtained by each of the multiplex systems that have been used.

2. Materials and methods

2.1. Samples

A total of approximately 800 DNA samples from unrelated males from the Iberian Peninsula were taken from collections of the authors. The samples can be regionally subdivided as follows: Aragón (33 individuals), Andalusia (168), Basque Country (116), Catalonia (92), Castilla-La Mancha (57), Castilla-León (129), Extremadura (52), Galicia (88), others/unknown (28). Four individuals showed apparent allele duplications (two samples showed three bands for DYS385, and two other samples showed two bands for DYS389, or DYS436) and some markers did not amplify in some samples; all these samples were excluded from further analysis. We also found one male who failed to give the Y-specific product from the amelogenin gene, which is consistent with his having a deletion of the *AMELY* gene [11], and one male with a DYS19 allele containing an apparent incomplete repeat. These anomalous individuals will be described in more detail in a future study. The final number of Iberian males with complete 19-locus Y-STR haplotypes was 763. General properties of haplotypes are described here; for further details, readers are asked to contact the authors.

2.2. EBF multiplex design and testing

Multiplex EBF (see electropherograms in Fig. 1) contains primers for seven Y-specific tetranucleotides and for amelogenin, the most widely used PCR-based sex-typing assay, which differentiates a 6 bp deletion in the X-linked copy of the amelogenin gene with respect to the Y-linked copy. The primers included in this multiplex were originally chosen from the literature and redesigned where necessary in order to ensure that they generated PCR products smaller than 300 bp, they did not form primer-dimers and they had similar annealing temperatures by using the programs Primer 3 (http://www2.no.embnet.org/primer3/primer3_www.cgi) and Amplify [12]. Finally, those primers generating PCR products with similar ranges of sizes were labelled with different fluorescent dyes. Amplification was

tested using serial dilutions containing from 50 to 0.1 ng/ μ l of human DNA. Female DNA showed amplification only for the X-derived product from the amelogenin gene (106 bp).

2.3. STR polymorphism typing

Nineteen Y-specific STRs and the amelogenin locus were amplified and scored in three multiplex reactions (see Table 1): multiplex MS1 was slightly modified from Thomas et al. [13], multiplex CTS is that described in Ayub et al. [6] and multiplex EBF is reported here for the first time. Physical locations of the microsatellites on the Y chromosome are shown in Fig. 2.

PCR reactions were performed in a 10 μ l final volume containing 200 μ M dNTPs, 10 mM Tris-HCL (pH 9.0), 0.1% (v/v) triton X-100, 0.01% (w/v) gelatin, 50 mM KCl, 2.2 mM MgCl₂, 0.26 U SuperTaq (HT Biotechnology Ltd.), 18.6 nM TaqStart monoclonal antibody (MAb, Clontech), a volume of 10 \times stock of primers and 5–20 ng genomic DNA. Primer sequences and concentrations used are indicated in Table 1.

PCR cycling conditions for multiplex MS1 were modified from Thomas et al. [13] as follows: 95 °C for 5 min; five cycles of 94 °C for 1 min, 57.5 °C \rightarrow 55 °C (–0.5 °C/cycle) for 1 min, 72 °C for 1 min; 30 cycles of 94 °C for 1 min, 55 °C for 1 min, 60 °C for 1 min; final extension of 72 °C for 10 min. PCR cycling conditions for multiplex CTS were as described in Ayub et al. [6] and those for multiplex EBF were: 95 °C for 5 min; 10 cycles of 94 °C for 1 min, 60 °C \rightarrow 55.5 °C (–0.5 °C/cycle) for 1 min, 72 °C for 1 min; 25 cycles of 94 °C for 1 min, 55 °C for 1 min, 60 °C for 1 min; and a final extension of 72 °C for 10 min.

Diluted PCR products were mixed with the internal lane standard ABI GS500 TAMRA and run in a 6% polyacrylamide gel using an ABI377 DNA sequencer. Genescan Version 3.0 and Genotyper Version 2.5 software packages were used to collect the data and analyse fragment sizes. Y chromosome STR alleles were named according to the number of repeat units they contain, which was established through the use of sequenced allelic ladders and/or reference DNA samples from collections of the authors, or provided by either Lutz Roewer or Chris Tyler-Smith. A quality assurance test offered by the Y-STR Haplotype Reference Database involving 'blind' haplotyping of five DNA samples (for loci DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 and DYS385) according to the ISFG guidelines for STR analysis [14] was successfully completed.

2.4. Nomenclature

Two previously described loci, YA7.1 and YA7.2 [5] have been assigned *D* segment numbers in accordance with ISFG guidelines, and are here referred to as DYS460 and DYS461, respectively. A new locus, identified from the Genome Database (GDB:6449613; <http://gdbwww.gdb.org/>) has the *D* segment number DYS462.

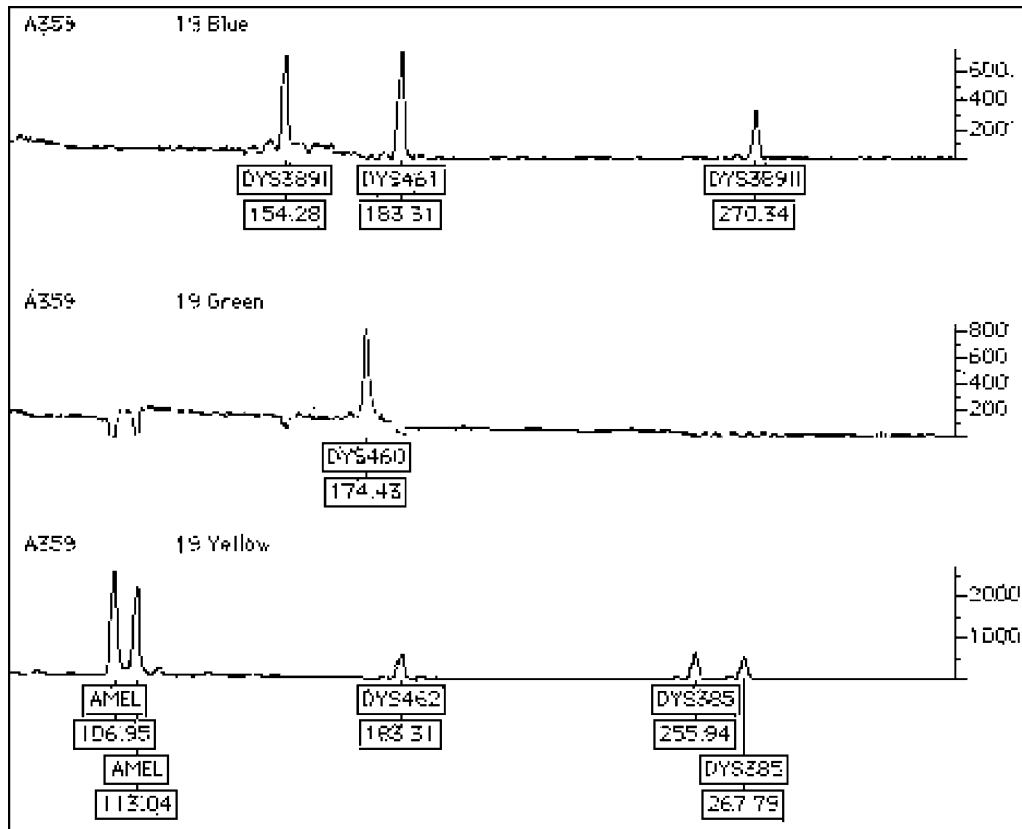


Fig. 1. Electropherograms for multiplex EBF. Electropherograms for sample A359: numbers below locus names indicate size in base pairs, as measured in an ABI377. Refer to Section 2 and Table 1 for multiplex details.

Table 1

Loci amplified in each multiplex, the repeat unit each contains, primer sequences, dyes and concentrations used

Locus	Repeat unit	Dye	Final (μM)	Forward (5'–3')	Reverse (5'–3')	Reference
MS1 multiplex						
DYS19	TAGA	TET	0.35	ACTGAGTTTCTGTTATAGTGTTTTT	ATGGCATGTAGTGAGGACA	New design/[4]
DYS388	ATT	TET	0.25	GTGAGTTAGCCGTTTAGCGA	CAGATCGCAACCACTGCG	[4]
DYS390	TCTG/A	FAM	0.20	TATATTTTACACATTTTGGGCC	TGACAGTAAAATGAACACATTGC	[4]
DYS391	TCTA	FAM	0.23	CTATTCATCAATCATACCCCATAT	ACATAGCCAAATATCTCCTGGG	[13]
DYS392	TAT	HEX	0.35	AAAAGCCAAAGAAGGAAAACAAA	CAGTCAAAGTGGAAAGTAGTCTGG	[13]
DYS393	AGAT	HEX	0.15	GTGGTCTTCTACTTGTGTCAATAC	AACTCAAGTCCAAAAAATGAGG	[4]
CTS multiplex						
DYS434	CTAT/TAAT	TET	0.2	CACTCCCTGAGTGCTGGATT	GGAGATGAATGAATGGATGGA	[6]
DYS435	TGGA	TET	0.05	AGCATCTCCACACAGCACAC	TTCTCTCTCCCCTCTCTC	[6]
DYS436	GTT	FAM	0.1	CCAGGAGAGCACACACAAAA	GCAATCCAACCTCAGCCAAT	[6]
DYS437	TCTA/G	HEX	0.1	GACTATGGGCGTGAGTGCAT	AGACCCTGTCAATCACAGATGA	[6]
DYS438	TTTTC/A	HEX	0.25	TGGGGAATAGTTGAACGGTAA	GTGGCAGACGCCTATAATCC	[6]
DYS439	GATA	TET	0.2	TCCTGAATGGTACTTCTTAGGTTT	GCCTGGCTTGGAAATCTTTT	[6]
EBF multiplex						
DYS385	GAAA	HEX	0.20	AGCATGGGTGACAGAGCTA	CCAATTACATAGTCTCTCTTC	[4]/[19]
DYS389	TCTG/A	FAM	0.35	CCAACTCTCATCTGTATTATCTATGT	CCTGAGTAGCAGAAGAATGTCATA	New design/[13]
DYS460	ATAG	TET	0.35	GCCAAACTCTTTCCAAGAAG	TCATCTATCCTCTGCCTATCATT	New design
DYS461	T/CAGA	FAM	0.25	AGGCAGAGGATAGATGATATGGAT	TTCAGGTAAATCTGTCCAGTAGTGA	[5]
DYS462	TATG	HEX	0.3	TGTGCTGTACCAGTTGCCTA	CCAGCCTGAGCAAGAGAGTA	New design
Amelogenin		HEX	0.15	CCCTGGGCTCTGTAAAGAATAGTG	ATCAGAGCTTAAACTGGGAAGCTG	[20]

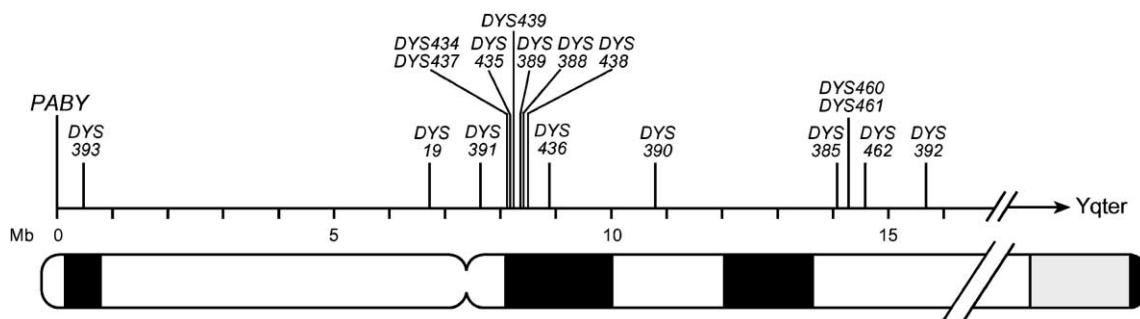


Fig. 2. Physical locations of the 19 microsatellites on the Y chromosome. Approximate positions are shown on a scale (in Mb) with reference to the pseudoautosomal boundary (PABY), and were estimated by placing microsatellite sequences within contigs (<http://www.ncbi.nlm.nih.gov/genome/seq/chr.cgi?CHR=Y>). Contigs and some individual clones were abutted using published mapping information [18], and three gaps within the region of interest ignored. Note that the order of DYS434/DYS437, DYS435 and DYS439 are reversed with respect to a previously published map [6]. Positions of G-bands on the chromosomal ideogram are schematic only.

Allele nomenclature was as follows: DYS19, DYS390, DYS392, DYS393, DYS389, DYS385 (YHRD); DYS388 [15]; and DYS435, DYS436, DYS438, DYS439 [6]. For DYS434, we follow a newly proposed nomenclature [16]

which, as well as the block of CTAT repeats, considers a flanking repeat unit of TAAT which can exist in one [8] or two [9,16] copies. For DYS437 we follow recent nomenclature [8,9,16]. Following the ISFG guidelines [14] we have

a) DYS460, 11 repeat allele, 173 bp

```
121 cttcagcaga aagagcccaa atttgccaaa ctctttccaa qaagaattat ctaggaaagt
181 caagacagta gcaagcacia gaataccaga ggaatctgac acctctgacA TAGATAGATA
241 GATAGATAGA TAGATAGATA GATAGATAGA TAGAtaatag acaaatacat aataaatgat
301 aggcagagga tagatga
```

b) DYS461, 12 repeat allele, 182 bp

```
301 aggcagagga tagatgatat ggatagacag atatatctaa taggtagatg atagataata
361 ggtagataga agataggTAG ATAGATAGAT AGATAGATAG ATAGATAGAT AGATAGATAG
421 ACAGAtaaga gagaaacaga aatatagtga cacagcatca ctactggaca gatttacctg
481 aa
```

c) DYS462, 13 repeat allele, 190 bp

```
83 tgtgctgtac cagttgccta tgagaattgc ttgtgacaac ttaggcctgt tccaggctaa
143 aatttttcca aatttttgta aaccatccca ggctacattt TATGTATGTA TGTATGTATG
203 TATGTATGTA TGTATGTATG TATGTATGTA TGTattttaga gacagggttt TACTCTCTTG
263 CTCAGGCTGG
```

Fig. 3. Allele sequences of loci DYS460, DYS461 and DYS462. (a) Sequence of locus DYS460. Original primers [5] are shown in bold lower-case, and primers used in this study are underlined. The repeat unit (upper-case) is here considered as ATAG [16], whereas previously it was GATA [5]. The initial G and final A of the original GATA repeat block is underlined and bold. Note that the redesignation of repeat unit does not change the number of repeats within the block. (b) Sequence of locus DYS461. Primers are as previously described [5], and are underlined. The repeat unit (upper-case) is here considered as TAGA [16], whereas previously it was GATA [5], and we also include the CAGA repeat following the array [16]. The initial G and final A of the original GATA repeat block are underlined and bold. Note that the redesignation of repeat units increases the number of repeats within the block by two. Note that Genbank entries for DYS460 (G42675) and DYS461 (G42671) are overlapping. Nucleotides in (b) and (c) are both labelled according to the G42675 entry so that the relationship between the two loci can be seen. (c) Sequence of locus DYS462 (this study). The sequence differs from the GDB entry (G09411) at five positions (not shown). Primer binding sites are underlined.

considered the repeat blocks for loci DYS460 and DYS461 to be ATAG and TAGA, respectively and not GATA as previously described [5]; details are in Fig. 3. Allele designation for the newly described locus DYS462 is also given in Fig. 3.

2.5. Statistical analysis

Allele frequencies for each locus were calculated by simple gene counting. DYS389 results are reported as DYS389I and DYS389AB, where DYS389AB [17] is the result of subtracting the smaller fragment (DYS389I) from the longer one (DYS389II). DYS385 represents variation at two loci simultaneously, which cannot be differentiated, and it is, therefore, analysed as a ‘phenotype’. Therefore, the 18 primer pairs used amplify 19 different STRs and the amelogenin locus.

Gene diversity (D , which corresponds to expected heterozygosity for autosomal loci) was computed for each locus using the formula: $D = (n/n - 1)(1 - \sum p_i^2)$, where n is sample size and p_i the allelic frequency. Haplotype diversity was computed with the same equation using haplotype frequencies instead of allele frequencies. For the Y chromosome, haplotype diversity is numerically identical to two parameters used in forensic genetics, i.e. PD and chance of exclusion (CE). Conversely, the probability of obtaining an identical haplotype in a pair of random unrelated males can be estimated as $1-D$.

3. Results and discussion

3.1. Single locus analysis

Allele frequency distributions for loci DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS389I/AB, DYS385 (phenotype frequencies), DYS434, DYS435, DYS436, DYS437, DYS438, DYS439, DYS460, DYS461 and DYS462 in the Iberian Peninsula are presented in the Appendix A.

With the exception of DYS392 and DYS438, all Y-STR loci typed showed a unimodal distribution (see Appendix A). The number of alleles found at each locus ranged from 3 to 8 with the exception of DYS385, which presented 45 phenotypes. The allele length variance ranged from 0.01 (DYS435) to 1.29 (DYS438) corresponding to the loci with the minimum and maximum number of alleles observed, respectively. Gene diversity values for each Y chromosome STR (Table 2) ranged from 0.0208 (DYS435) to 0.7997 (DYS385). As discussed below, the most diverse loci were the DYS385 system and DYS439.

3.2. Haplotype analysis

Loci were combined to produce haplotypes with the loci in the order DYS19-DYS388-DYS389I-DYS389AB-

Table 2
Cumulative distribution of haplotype diversity^a

Locus	Gene diversity	Cumulative number of haplotypes	Cumulative haplotype diversity
DYS385	0.7997	45	0.7997
DYS439	0.6683	116	0.9216
DYS391	0.5940	180	0.9592
DYS390	0.5733	271	0.9766
DYS389I	0.5630	344	0.9883
DYS460	0.5470	430	0.9944
DYS437	0.5430	480	0.9960
DYS438	0.5400	501	0.9965
DYS19	0.5157	533	0.9971
DYS389AB	0.5074	574	0.9978
DYS392	0.5014	585	0.9979
DYS461	0.4839	610	0.9985
DYS462	0.4125	621	0.9985
DYS393	0.3809	630	0.9988
DYS388	0.3343	634	0.9988
DYS434	0.0489	635	0.9988
DYS436	0.0209	635	0.9988
DYS435	0.0208	635	0.9988

^a The cumulative number of haplotypes and haplotype diversity indicate the number of different haplotypes and the haplotype diversity that would be obtained by considering the loci from the top of the table to that particular locus, respectively.

DYS390-DYS391-DYS392-DYS393-DYS385-DYS434-DYS435-DYS436-DYS437-DYS438-DYS439-DYS460-DYS461-DYS462. Complete haplotypes were available for 763 individuals, which presented 635 different haplotypes. The most frequent haplotype (Table 3) was 14-12-13-16-24-11-13-13-(11, 14)-11-11-12-15-12-12-11-12-11, which contains the most frequent allele at each locus; it was found in 16 individuals (2.09%). The two next most frequent haplotypes were found in 8 and 7 individuals; however, the number of haplotypes shared fell progressively (see Table 3): two haplotypes were shared by 6 individuals, three by 5, 6 by four, 12 by three, 36 by two and 573 of 635 haplotypes (90.24%) were found in single individuals.

Haplotype gene diversity using all loci was 0.9988. This means that in the overall Iberian population, two unrelated males picked at random will on average carry the same 19-locus STR haplotype in only 0.12% (1/833) of the cases.

3.3. Forensic usefulness of the new markers

Loci were ranked according to their diversity (Table 2) and the cumulative number of haplotypes and haplotype diversity value computed by considering the loci from the top (the most variable locus) of the table to each particular locus. The table outlines in decreasing order which are the most informative loci and how many loci are required to

Table 3
Fourteen haplotypes found in more than three individuals

Absolute frequency	% ^a	DYS19	DYS388	DYS389I	DYS389AB	DYS390	DYS391	DYS392	DYS393	DYS385	DYS434	DYS435	DYS436	DYS437	DYS438	DYS439	DYS460	DYS461	DYS462
16	2.09	14	12	13	16	24	11	13	13	11, 14	11	11	12	15	12	12	11	12	11
8	1.05	14	12	13	16	24	10	13	13	11, 14	11	11	12	14	12	12	10	12	11
7	0.92	14	12	13	16	24	10	13	13	11, 14	11	11	12	15	12	12	11	12	11
6	0.79	14	12	13	16	24	11	13	13	11, 14	11	11	12	14	12	12	10	12	11
6	0.79	14	12	14	16	24	11	13	13	11, 14	11	11	12	15	12	11	11	12	11
5	0.66	14	12	13	16	24	11	13	13	11, 14	11	11	12	15	12	13	11	12	11
5	0.66	14	12	13	16	24	11	13	13	12, 14	11	11	12	15	12	11	11	12	11
5	0.66	14	12	12	16	24	11	13	13	11, 14	11	11	12	15	12	12	11	12	11
4	0.52	14	12	13	16	24	11	13	13	11, 14	11	11	12	15	12	12	10	12	11
4	0.52	14	12	13	16	24	11	13	13	11, 14	11	11	12	15	12	13	10	12	11
4	0.52	14	12	13	16	24	11	13	13	11, 14	11	11	12	15	12	11	11	12	11
4	0.52	14	12	14	16	23	10	13	13	11, 11	11	11	12	15	12	13	11	12	11
4	0.52	14	12	14	16	24	11	13	13	11, 14	11	11	12	15	12	13	10	12	11
4	0.52	16	13	13	15	23	10	11	13	12, 12	11	11	12	15	10	11	10	11	12

^a % is the percentage in 763 individuals.

Table 4
Incremental haplotype diversity contributed by additional loci to the 9-locus 'core' haplotype^a

Locus	Cumulative number of haplotypes	Cumulative haplotype diversity
9-Locus 'core' haplotype	417	0.9896
DYS439	504	0.9954
DYS460	567	0.9977
DYS437	589	0.9983
DYS438	602	0.9984
DYS461	621	0.9987
DYS462	630	0.9988
DYS388	634	0.9988
DYS434	635	0.9988
DYS436	635	0.9988
DYS435	635	0.9988

^a The cumulative number of haplotypes and haplotype diversity indicate the number of different haplotypes and the haplotype diversity that would be obtained by considering the loci from the top of the table to that particular locus, respectively. The 9-locus 'core' haplotype contains DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393.

achieve a particular PD. The additional haplotype diversity achieved by adding extra loci to the 9-locus 'core' haplotype used for the Y-STR Haplotype Reference Database is shown in Table 4.

In the Iberian Peninsula, DYS434, DYS435 and DYS436 showed the lowest discrimination values (Tables 2 and 4). The same result was obtained in populations from Pakistan [6,7] but not in Chinese populations where loci DYS434 and DYS435 showed higher discrimination values [8]. Therefore, although these are not likely to be the markers of choice for forensic work involving individuals of European ancestry, and might be omitted from forensic analyses where cost is a concern, they may be informative for identifying other continental origins. Gusmão et al. [9] studied a subset (DYS434, DYS437, DYS438 and DYS439) of the loci used here in populations from north Portugal, Mozambique and Macao and found very low gene diversity values for DYS434 in both north Portugal and Mozambique and for DYS437 in Mozambique.

In contrast, DYS439, DYS460, DYS437 and DYS438 yielded high diversity values in the Iberian Peninsula, equivalent to the loci commonly used by the forensic community. This result suggests that these loci, and maybe DYS461 and DYS462, should be considered in order to improve the exclusion probabilities in forensic cases. When these six loci are included in addition to the standard nine loci, the number of different haplotypes increases from 417 to 630, and the haplotype diversity increases from 0.9896 to 0.9988. Despite this, the increment in number of haplotypes is clearly less spectacular beyond the 12 most informative loci. Most haplotypes (630 of 635) were identified using the 14 most variable loci of the 19

Table 5
Number of haplotypes, haplotype diversity and chance of matching obtained by each multiplex

System	Number of haplotypes	Haplotype diversity	Chance of matching
MS1 multiplex	198	0.9292	0.0708 (1/14)
EBF multiplex	347	0.9812	0.0188 (1/53)
CTS multiplex	83	0.9089	0.0911 (1/11)
All loci	635	0.9988	0.0012 (1/833)

studied. Moreover, haplotype diversity beyond this point remained constant.

When comparing the three multiplex systems used (Table 5) EBF was the most informative generating 347 different haplotypes with a haplotype diversity (0.9812) clearly higher than that of the other multiplex reactions used. This is due to the inclusion of DYS385, which is by far the most informative system. When used in addition to the most commonly used set of Y-STR loci, this multiplex would add the value of the new loci DYS460, DYS461 and DYS462. When we do not obtain amplification for any of the Y-specific STR loci included in the EBF multiplex the inclusion of the amelogenin locus confirms whether that is due to the sex of the individual providing the sample or to the absence of amplifiable DNA. Furthermore, this multiplex system also allows the detection of those males that lack the Y copy of the amelogenin gene as a result of an interstitial deletion event, which have been described at low frequencies but may be of importance in certain circumstances [11]. Finally, although multiplex CTS generates the lowest haplotype diversity value, it contains some informative loci such as DYS439, DYS437 and DYS438 and, therefore, it should also be considered when it is necessary to increase the discriminative power obtained by the commonly used Y-STR loci.

Knowledge about the geographical distribution of Y-STR haplotypes and informativeness of the Y-specific STRs used to construct them is crucial for the forensic community. Further characterisation of the newly emerging loci and the establishment of a unique nomenclature are desirable objectives.

Acknowledgements

E.B. and A.C.L. are supported by the Wellcome Trust. M.A.J. is a Wellcome Trust Senior Fellow in Basic Biomedical Science (Grant number 057559). This work has been financed in part by Complutense University Project PR48/01-9837. Subsets of samples were kindly provided by J. Bertranpetit, A. Carracedo and L. Fañanas. We thank Connie Talbot of GDB for assistance with *D* segment numbers and L. Roewer and C. Tyler-Smith for control DMAs.

Appendix A

Allele frequencies of 19 STR loci in 763 Iberian males

	DYS19 ^a	DYS388	DYS389I	DYS389AB	DYS390	DYS391	DYS392	DYS393	DYS434	DYS435	DYS436	DYS437	DYS438	DYS439	DYS460	DYS461	DYS462
7													0.004				
8													0.003	0.001	0.003	0.001	0.003
9					0.084						0.001		0.087	0.003	0.018		
10			0.001		0.413			0.008	0.001				0.201	0.064	0.408	0.017	0.024
11	s	0.004	0.003		0.478	0.275	0.001	0.975	0.990	0.004			0.041	0.295	0.535	0.143	0.734
12	0.008	0.810	0.139		0.022	0.037	0.123	0.017	0.009	0.990			0.641	0.467	0.035	0.691	0.220
13	0.104	0.080	0.596		0.003	0.649	0.772			0.004	0.003		0.024	0.153	0.001	0.138	0.018
14	0.667	0.028	0.250	0.003		0.035	0.092			0.001	0.341		0.001	0.017		0.009	0.001
>14	0.001																
15	0.166	0.046	0.010	0.076		0.004	0.012					0.579				0.001	
16	0.029	0.029		0.666								0.077					
17	0.025	0.004		0.206													
18				0.042													
19				0.007													
20				0.001													
21					0.005												
22					0.048												
23					0.248												
24					0.595												
25					0.098												
26					0.005												

^a For DYS19, '>14' indicates an allele assumed to contain fourteen repeats plus an incomplete repeat unit.

Phenotype frequencies for DYS385 in 763 Iberian males

Phenotype	DYS385
9, 14	0.001
10, 13	0.003
10, 14	0.008
10, 15	0.005
11, 11	0.022
11, 12	0.005
11, 13	0.050
11, 14	0.425
11, 15	0.087
11, 16	0.013
11, 17	0.004
12, 12	0.028
12, 13	0.008
12, 14	0.060
12, 15	0.020
12, 16	0.010
12, 17	0.004
12, 18	0.005
13, 13	0.004
13, 14	0.046
13, 15	0.024
13, 16	0.021
13, 17	0.008
13, 18	0.009
13, 20	0.001
13, 21	0.001
14, 14	0.034
14, 15	0.014
14, 16	0.010
14, 17	0.003
14, 18	0.004
15, 15	0.012
15, 16	0.008
15, 17	0.004
15, 18	0.003
15, 19	0.001
16, 16	0.003
16, 17	0.008
16, 18	0.012
16, 19	0.001
17, 17	0.003
17, 18	0.001
17, 19	0.003
18, 18	0.004
18, 19	0.001

References

- [1] M.A. Jobling, A. Pandya, C. Tyler-Smith, The Y chromosome in forensic analysis and paternity testing, *Int. J. Legal Med.* 110 (1997) 118–124.
- [2] M.A. Jobling, N. Bouzekri, P.G. Taylor, Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1), *Hum. Mol. Genet.* 7 (1998) 643–653.
- [3] P. de Knijff, M. Kayser, A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Pérez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, L. Roewer, Chromosome Y microsatellites: population genetic and evolutionary aspects, *Int. J. Legal Med.* 110 (1997) 134–140.
- [4] M. Kayser, A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Pérez-Lezaun, A. Piccinini, 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.
- [5] P.S. White, O.L. Tatum, L.L. Deaven, J.L. Longmire, New, male-specific microsatellite markers from the human Y chromosome, *Genomics* 57 (1999) 433–437.
- [6] Q. Ayub, A. Mohyuddin, R. Qamar, K. Mazahar, T. Zerjal, S.Q. Mehdi, C. Tyler-Smith, Identification and characterisation of novel human Y chromosomal microsatellites from sequence database information, *Nucl. Acids Res.* 28 (2000) e8.
- [7] A. Mohyuddin, Q. Ayub, R. Qamar, T. Zerjal, A. Helgason, S.Q. Mehdi, C. Tyler-Smith, Y chromosomal STR haplotypes in Pakistani populations, *Foren. Sci. Int.* 118 (2001) 141–146.
- [8] Y.P. Hou, J. Zhang, Y.B. Li, J. Wu, S.Z. Zhang, M. Prinz, Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han populations, *Foren. Sci. Int.* 118 (2001) 147–152.
- [9] L. Gusmão, C. Alves, A. Amorim, Molecular characterisation of four human Y-specific microsatellites (DYS434, DYS437, DYS438 DYS439) for population and forensic studies, *Ann. Hum. Genet.* (2001) 285–291.
- [10] L. Roewer, M. Krawczak, S. Willuweit, M. Nagy, C. Alves, A. Amorim, K. Anslinger, C. Augustin, A. Betz, E. Bosch, A. Caglià, A. Carracedo, D. Corach, T. Dobosz, B.M. Dupuy, S. Füredi, C. Gehrig, L. Gusmão, J. Henke, L. Henke, M. Hidding, C. Hohoff, B. Hoste, M.A. Jobling, H.J. Kärger, P. de Knijff, R. Lessig, E. Liebeherr, M. Lorente, B. Martínez-Jarreta, P. Nievas, M. Nowak, W. Parson, V.L. Pascali, G. Penacino, R. Ploski, B. Rolf, A. Sala, U. Schmidt, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, M. Kayser, Online reference database of Y-chromosomal short tandem repeat (STR) haplotypes, *Foren. Sci. Int.* 118 (2001) 106–113.
- [11] F.R. Santos, A. Pandya, C. Tyler-Smith, Reliability of DNA-based sex tests, *Nat. Genet.* 18 (1998) 103.
- [12] W. Engels, Amplify—For analysing PCR Experiments, University of Wisconsin, Madison, WI, 1992.
- [13] M.G. Thomas, N. Bradman, H.M. Flinn, High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y chromosome, *Hum. Genet.* 6 (1999) 577–581.

- [14] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, M. Krawczak, W.R. Mayr, N. Morling, B. Olaisen, V. Pascali, M. Prinz, L. Roewer, P.M. Schneider, A. Sajantila, C. Tyler-Smith, DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, *Int. J. Legal Med.* 114 (2001) 305–309.
- [15] A. Pérez-Lezaun, F. Calafell, D. Comas, E. Mateu, E. Bosch, R. Martínez-Arias, J. Clarimón, G. Fiori, D. Luiselli, F. Facchini, D. Pettener, J. Bertranpetit, Sex-specific migration patterns in central Asian populations, revealed by analysis of Y chromosome short tandem repeats and mtDNA, *Am. J. Hum. Genet.* 65 (1999) 208–219.
- [16] L. Gusmão, A. González-Neira, C. Alves, M. Lareu, S. Costa, A. Amorim, A. Carracedo, Chimpanzee homologs of human Y-specific STRs: a comparative study and consequences for nomenclature, *Foren. Sci. Int.*, in press.
- [17] B. Rolf, E. Meyer, B. Brinkmann, P. de Knijff, Polymorphism at the tetranucleotide repeat locus DYS389 in 10 populations reveals strong geographic clustering, *Eur. J. Hum. Genet.* 6 (1998) 583–588.
- [18] C.A. Tilford, T. Kuroda-Kawaguchi, H. Skaletsky, S. Rozen, L.G. Brown, M. Rosenberg, J.D. McPherson, K. Wylie, M. Sekhon, T.A. Kucaba, R.H. Waterston, D.C. Page, A physical map of the human Y chromosome, *Nature* 409 (2001) 943–945.
- [19] P.M. Schneider, S. Meuser, W. Waiyawuth, Y. Seo, C. Rittner, Tandem repeat structure of the duplicated Y chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations, *Foren. Sci. Int.* (1998) 61–70.
- [20] K.M. Sullivan, A. Mannucci, C.P. Kimpton, P. Gill, A rapid and quantitative DNA sex test—fluorescence-based PCR analysis of X–Y homologous gene amelogenin, *Biotechniques* 15 (1993) 636–641.