TECHNICAL NOTE

Florence Rousselet, M.S., Helene Pfitzinger, Ph.D.; and Patrice Mangin, M.D., Ph.D.

A Pentaplex Automated Fluorescent Typing System for Forensic Identification and French Caucasian Population Data*

REFERENCE: Rousselet F, Pfitzinger H, Mangin P. A pentaplex automated fluorescent typing system for forensic identification and french caucasian population data. J Forensic Sci 1997; 42(3):500–503.

ABSTRACT: The polymerase chain reaction (PCR) amplification of short tandem repeat (STR) loci has already proven to be a method of choice for large scale typing of DNA samples in which the conventional restriction fragment length polymorphism (RFLP) technique is ineffective.

A quadruplex PCR including HUMvWFA31A, HUMF13A01, HUMTH01, and HUMFESFPS STR loci is used successfully for routine forensic applications in our laboratory. However, the need to increase the discrimination power of the PCR systems used prompted us to develop a second system of a pentaplex PCR for the analysis of 4 additional STR loci (HUMD8S1179, HUMD18S51, HUMD21S11, and HUMFIBRA) and the sex determination by amplification of a segment of the X-Y homologous Amelogenin gene. Allele and phenotype frequencies for these 4 STR systems were obtained by multiplex amplification, from approximately 200 randomly selected and unrelated French Caucasian individuals. Statistical calculations for these phenotype distributions met expectations for Hardy-Weinberg equilibrium. Furthermore, the French allelic frequencies of D18S51, D21S11, and HUMFIBRA loci were compared with the data obtained by the Forensic Science Service (UK) for the British Caucasian population and proved to be similar.

KEYWORDS: forensic science, DNA typing, population data, short tandem repeat, multiplex polymerase chain reaction, fluorescence, france, D8S1179, D18S51, D21S11, FIBRA, Amelogenin

The analysis of STR (short tandem repeat) loci by PCR (polymerase chain reaction) is a very valuable method for the forensic identification of small and highly degraded DNA specimens. In fact, the application of multiplex PCR utilizing 4 different STR loci such as HUMvWFA31A, HUMF13A01, HUMTH01, and HUMFESFPS has been shown to be efficient for human individual identification (1).

The repeated DNA fragments found in STR alleles exhibit three main types of sequence structure. The first type is made of simple repeats containing identical units (e.g., HUMFESFPS, HUMTH01, and HUMF13A01), the second type consists in compound repeats containing several different simple repeats (e.g., HUMvWFA31A) and the third type concerns complex repeats containing several units of variable length (e.g., HUMD21S11). Furthermore, some highly polymorphic STR loci consisting of imperfect tetranuleotide repeats show intercalary alleles differing in size by only 2 bp. HUMD18S51 (D18S51), HUMD21S11 (D21S11), and HUMFIBRA (FGA) loci used in our pentaplex PCR are such imperfect STRs. HUMD8S1179 (D8S1179), D18S51, D21S11, and FGA STR loci already proved to coamplify well with other STR loci and with loci such as the amelogenin gene (2). The amelogenin-linked sex test was included in our pentaplex system as the sex determination of biological samples can be of real interest in some forensic cases (3).

The 4 STRs studied in this work exhibit between 11 and 18 observed alleles per locus and have a combined matching probability of 1.5 by 10^{-10} when associated with the four previous STR loci.

The conditions for the pentaplex amplification of the D8S1179, D18S51, D21S11, and FGA loci with the amelogenin locus are presented, as well as the French Caucasian allelic frequencies. Comparisons for allelic distributions were conducted with the published British Caucasian data (2).

Materials and Methods

DNA Extraction and Purification

The DNA extraction on blood and tissues, by the phenol-chloroform protocol, has been described previously (4).

Loci and Primer Sequences

STR Loci—The primer sequences for the five loci have already been described in the literature (5). The 4 STR loci analyzed are located on different chromosomes: Chromosome 4, 8, 18, and 21 for FGA (6), D8S1179 (5), D18S51 (7), and D21S11 (8) loci, respectively.nb

The D21S11 locus has a complex structure (TCTR tetranucleotide repeat motif, where R may be an A or a G) with the occurrence of microheterogeneities and is the most studied locus among the 4 STR loci presented in this study. Sequence analysis of D21S11 alleles has revealed a 2-bp mutation within the repeat region, accounting for the 2-bp allele differences observed (9). Furthermore, the D21S11 allele designation is based on the total number of "TV" dinucleotide repeats (with V being an A, a C, or a G) present in each amplification product (10). For FGA and D18S51 loci, alleles are designated arbitrarily (as their exact sequences are not known) (2) and alleles with 2-bp differences are designated as ".2".

Institut de Médecine Légale, 11 Rue Humann, 67085 Strasbourg Cedex, France.

*Supported by a grant from the French ANRT (Association Nationale de la Recherche Technique).

Received April 1996; and in revised form 1 July 1996; accepted 3 Oct. 1996.

The X-Y Homologous Amelogenin Gene (Amel)

The amelogenin sex-typing system is based on the amplification of the X chromosome amelogenin gene (AMG) and the amelogenin-like sequence (AMGL) located near the centromere of the Y chromosome (11). A single set of primers flanking the 6-bp deletion within intron 1 of the X homologue were designed to result in short PCR products (106 and 112 bp for X and Y chromosomes respectively), allowing coamplification with microsatellite loci. This sex-typing system is very sensitive as it may detect as little as one cell (12) or severely degraded DNA. Furthermore, 1 ng of male DNA may be detected in a 100 fold excess of female DNA (3). This PCR test can also be a quantitative test as there is an equal copy number on both chromosomes, thus permitting the detection of aneuploidy such as XYY and XXY individuals or cells (13). The X specific fragment also acts as an internal control preventing mistyping due to lack of Y product. Another interesting property of the Amelogenin sex-test is the limitation of allelic drop-out, as the PCR products are approximately of the same size (3).

PCR Amplification and Fluorescent Detection

In order to distinguish the overlapping allele size ranges of the 4 STR loci analyzed at once, one primer per locus was labeled with a fluorescent dye marker (Genset-France). One Amel, D18S51 or D21S11 primer was labeled with the 6-FAM dye (6-carboxyfluorescein), one of the D8S1179 primers was labeled with the TET dye (Tetrachloro-6-carboxyfluorescein) and one of the FGA primers was labeled with the HEX dye (Hexachloro-6-carboxyfluorescein). The pentaplex PCR was performed in a final volume of 25 μL containing 5 to 10 ng of DNA isolated from whole blood, 1X Gibco-BRL buffer, 1.5 mM MgCl2, 0.05% Tween, 200 μg/mL BSA (Gibco-BRL), 200 μM each dNTP (MBI Fermentas), 0.625 U Taq DNA-polymerase (Gibco-BRL), 0.05 μM AmelA and AmelB primers, 0.1 μM each D18S51 primers, 0.25 μM each D21S11 primers, 0.4 μM each FGA primers and 0.8 μM each D8S1179 primers. The 32 PCR thermal cycles were carried out in a Perkin Elmer Gene Amp™ PCR system 9600, as follows: 30 s at 93°C, 75 s at 58°C, and 15 s at 72°C. The last cycle was followed by a 10 min extension period at 72°C.

An aliquot (2 μL) of each PCR product was loaded together with the Genescan-2500™ Tamra standard ladder on a 6% polyacrylamide denaturing gel and electrophoresed in 1X TBE buffer for 8 h at a constant power of 30 watts using the ABD (Applied Biosystems Division) 373 sequencer. The results were automatically analyzed by the ABD GENESCAN 672 software using the local Southern method. Final allele designation was determined by comparison of the size attributed to each sample to the size ascribed to the alleles in the allelic ladders which were run simultaneously on the same gel. FGA, D8S1179, D18S51, and D21S11 allelic ladders contain some of the alleles for each locus (alleles 8 to 17 for D8S1179, alleles 10 to 22 plus intercalary alleles 13.2 and 19.2 for D18S51, odd-numbered alleles from 59 to 77 plus intercalary alleles 54, 66, 68, 72, and 74 for D21S11 and alleles 17 to 29 plus intercalary alleles 18.2, 19.2, 20.2, 22.2, 23.2, 24.2, and 46.2 for FGA) and were provided by the Forensic Science Service (UK).

Statistical Evaluations

Observed and expected phenotypes were compared using the chi-square test to evaluate the Hardy-Weinberg (H-W) equilibrium: Phenotype classes with less than five expected events were pooled. The formula (c − 1 − a) were c is the number of compared phenotype classes and a the number of alleles appearing in these classes allowed us to determine the degree of freedom (df) of the test.

The French and English Caucasian population data were compared using a computer program from G. Carmody (Carleton University, Canada) which calculates the chi-square and G-statistics values, and the probability of getting random tables having larger values than the observed tables. A probability greater than 0.05 indicates that the population samples tested are homogeneous.

The matching probabilities (Pm) were calculated by the method of Jones (14).

The allelic diversity (h) is calculated according to the formula:

$$h = (1 - \sum x_i^2) \cdot (N/N - 1)$$

where x_i represents the allelic frequency of allele i, and N, the number of observed alleles (15).

Results and Discussion

The genotype data for 4 tetrameric STR loci were determined by multi-locus PCR reactions and fluorescence-based automated detection on a 373 ABD sequencer. The loci studied were chosen because they proved to coamplify reliably and to be highly polymorphic. Furthermore, the location of the loci on 4 different chromosomes eliminates the possibility of physical genetic linkage (5).

Pentaplex Amplification

The multiplex PCR reaction conditions described (see Materials and Methods) allowed us to obtain equilibrated amplifications at the different loci within the pentaplex. Our amplification conditions are different from those described by Oldroyd (5), except for the cycling parameters that were maintained. The PCR reaction volume was reduced to 25 μL and the primer concentrations were adjusted as a function of each primer batch concentration and quality. The primer concentrations needed were different from those optimized by Oldroyd et al., who analyzed the same 5 loci in the presence of three additional loci in an octoplex reaction and using different batches of primers.

The pentaplex PCR performed on blood DNA samples, permitted an efficient co-amplification of all five loci as shown on Fig. 1 and allowed the simultaneous development of the data bases for D8S1179, D18S51, D21S11, and FGA. Furthermore, both alleles of a heterozygote individual displayed similar fluorescent intensities within the same locus as both allele-peaks have equivalent sizes (Fig. 1). As expected, identical profiles were also obtained from DNA samples extracted from different organs (liver, spleen, kidney, lymph nodes, psoa muscles, and heart) of the same individual (data not shown). The pentaplex-PCR conditions were optimized for 5 ng of DNA template as for lower concentrations of DNA the amplifications at the different loci within the pentaplex were not equilibrated anymore (data not shown).

In all data base amplifications, the STR electrophoregrams were free of artifactual bands and did not show double peaks or stutter bands that are generally observed as a small peak preceding the main allele peak.

Population Data—Table 1 shows the observed allele frequencies for the 4 STR loci in the French Caucasian population sample.

HUMD8S1179 Locus—The 11 observed alleles shown in Table 1 range from 163 to 203 base pairs (bp). By analyzing 214 French

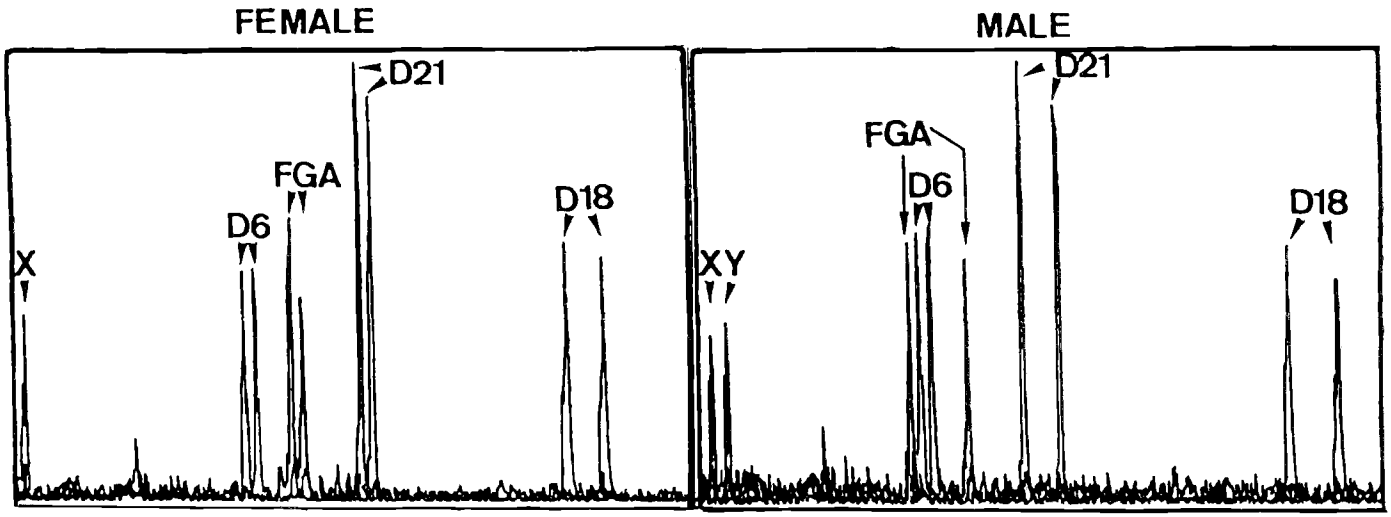

FIG. 1—*Electrophoregrams of pentaplex amplified blood DNA samples. The female individual (left) and the male individual (right) are both heterozygote at the 4 STR loci.*

TABLE 1—*Distributions of the observed alleles in a French Caucasian population sample, for D8S1179, D18S51, D21S11, and HUMFIBRA loci. The allele designation (ID) corresponds to the new nomenclature proposed by the Forensic Science Service (UK).*

D21S11 (N* = 232)		HUMFIBRA (N* = 232)		D18S51 (N* = 229)		D8S1179 (N* = 214)	
Allele ID	Frequency (%)	Allele ID	Frequency (%)	Allele ID	Frequency (%)	Allele ID	Frequency (%)
54	0.2	17	0.4	10	1.1	8	2.3
57	0.2	18	1.7	11	0.2	9	0.5
58	0.2	19	7.1	12	14.2	10	9.1
59	4.3	19.2	0.2	13	13.5	11	8.9
61	14.2	20	14.9	14	15.9	12	14
62	0.2	21	14.9	15	13.8	13	30.2
63	23.7	21.2	0.4	16	12.4	14	21.3
64	0.2	22	19.4	17	12	15	10.5
65	24.4	22.2	1.3	18	7.9	16	2.8
66	2.4	23	12.9	19	4.4	17	0.2
67	5.2	23.2	0.4	20	2.4	18	0.2
68	11.7	24	12.1	21	1.1		
69	1.7	24.2	0.2	22	0.9		
70	6.7	25	9.1	23	0.2		
71	0.4	26	3.7				
72	3.7	27	1.1				
74	0.4	46.2	0.2				
76	0.2						

*N represents the number of typed individuals.

Caucasian individuals, we found 3 major alleles (12, 13, and 14) and 3 very rare ones (9, 17, and 18). Concerning the phenotype distribution, we observed 33 out of the 66 possible phenotypes. The observed heterozygosity, allelic diversity and discrimination power for this locus are listed in Table 2. According to the chi-square test, no significant deviation from Hardy-Weinberg Equilibrium was found for our French Caucasian sample ($\chi^2 = 15.51$; df = 8; $0.3 < p < 0.5$).

TABLE 2—*Allelic diversity (h), observed heterozygosity, and discrimination power at D8S1179, D18S51, D21S11, and HUMFIBRA loci, for a French Caucasian population sample.*

	D8S1179	D18S51	D21S11	HUMFIBRA
Allelic Diversity (h)	0.82	0.88	0.84	0.87
Observed Heterozygosity	0.77	0.89	0.82	0.86
Power of Discrimination (pd)	0.94	0.97	0.95	0.97

HUMD18S51 Locus—Sixteen different possible alleles (9 to 24) which range from 272 to 332 bp have been identified for this locus. Alleles 9 and 24 were not observed in our French Caucasian sample of 229 individuals. Allele 11, the smallest allele observed, and allele 23, the largest one, are very rare. The six most frequent alleles (12, 13, 14, 15, 16, and 17) are nearly equally represented as shown in Table 1. Intercalary alleles (13.2 and 19.2), observed in Afro-Caribbeans and Asian populations (2), have not been found in our French Caucasian population. We observed 51 out of the 105 possible phenotypes. Our Caucasian sample did not deviate from Hardy-Weinberg Equilibrium ($\chi^2 = 21.01$; df = 12; $p > 0.05$). The observed heterozygosity, the allelic diversity, and the discrimination power of the D18S51 locus are indicated in Table 2.

HUMD21S11 Locus—The 26 D21S11 possible alleles (54 to 79) range from 207 to 257 bp. In our population sample of 232 individuals, the smallest allele observed migrated as allele 54 and the largest one migrated as allele 76. We did not observe alleles

55, 56, 60, 73, 75, 77, 78, and 79, and some other alleles (54, 57, 58, 62, 64, 71, 74, and 76) are very rare. The intercalary alleles, designated as even numbers, are often rare, except alleles 66, 68, 70, and 72 as shown in Table 1. Concerning the phenotype distribution, we observed 46 out of the 171 possible phenotypes. Table 2 indicates the observed heterozygosity, the allelic diversity, and the discrimination power of the locus. The chi-squared test performed on our French Caucasian data did not show any significant deviation from Hardy-Weinberg Equilibrium ($\chi^2 = 12.59$; df = 6; $0.5 < p < 0.9$).

HUMFIBRA Locus—The observed alleles (17 to 46.2) range from 175 to 293 bp. The smallest allele migrated as allele 17 and the largest one migrated as allele 46.2. To date, alleles 17.2, 18.2, 20.2, 25.2, 26.2, 27.2, 28, 28.2, and 29 to 46 have not been observed in the French Caucasian sample (Table 1). The French FGA phenotypes distribution is in Hardy-Weinberg Equilibrium ($\chi^2 = 21.03$; df = 12; $0.2 < p < 0.3$) and the observed heterozygosity, the allelic diversity and the discrimination power were calculated (Table 2).

Allelic diversity and observed heterozygosity were very close for each locus. The homogeneity G-statistics test performed for the D18, D21, and FGA loci, in order to compare the French and the English (2) Caucasian population samples led to the conclusion that the 2 samples were homogeneous ($G = 16.22$, $p = 0.25$ for D18S51; $G = 11.80$, $p = 0.87$ for D21S11 and $G = 18.70$, $p = 0.25$ for FGA).

Conclusion

The fluorescent detection of PCR amplified STRs is now frequently used for human identification in laboratories, as it is rapid, sensitive and highly discriminating. A first quadruplex-PCR system, including HUMTH01, HUMFESFPS, HUMvWFA31A, and HUMF13A01, already used for applications in our laboratory provides a matching probability of 7.4 by 10^{-5} for a randomly selected sample. The use of a second multi-locus PCR system as the pentaplex system described herein further increases the discrimination power obtainable through PCR analyses. Moreover, the use of imperfect tetranucleotide STRs, such as HUMFIBRA, D18S51, and D21S11, that exhibit a higher polymorphism owing to the occurrence of intercalary alleles and consequently lower matching probability values, reduces significantly the combined matching probability of the PCR-system. Furthermore the inclusion of the XY test in the multiplex STR-PCR can provide valuable information in forensic identification cases. The combination of the 2 multiplex-PCR systems utilizing 8 STR loci is discriminating enough for forensic identification. In fact, although the availability of four different fluorescent dyes and the ranges of sizes of the chosen STRs allow the inclusion of additional loci in each multiplex to further increase the discrimination, the matching probability for the combination of the 8 polymorphic STRs described in this study already reaches 1.5 by 10^{-10} (Table 3), a value that may be compared to the one obtained with RFLPs.

Acknowledgment

We wish to thank the Forensic Science Service (UK) for their advice and for providing D8S1179 primer sequences and all allelic ladders.

TABLE 3—*Matching probabilities (Pm) at 8 STR loci for a French Caucasian population sample. The amelogenin locus included in the pentaplex-PCR was not considered for the calculation of the matching probabilities.*

Locus	Pm	Locus	Pm
HUMTH01	0.086	HUMD8S1179	0.057
HUMFESFPS	0.153	HUMD18S51	0.027
HUMVWA31A	0.062	HUMD21S11	0.045
HUMF13A1	0.091	HUMFIBRA	0.030
QUADRUPLEX	7.4×10^{-5}	PENTAPLEX	2.1×10^{-6}
QUADRUPLEX AND PENTAPLEX (8 STRs combined)			1.5×10^{-10}

References

1. Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M. Automated DNA profiling employing multiplex amplification of Short Tandem Repeat loci. PCR Meth Appl 1993;3:13–22.
2. Urquhart A, Oldroyd NJ, Kimpton CP, Gill P. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. Biotechniques 1995;18:116–21.
3. Mannucci A, Sullivan KM, Ivanov PL, Gill P. Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. Int J Leg Med 1994;106:190–3.
4. Pfitzinger H, Ludes B, Mangin P. Sex determination of forensic samples: Co-amplification and simultaneous detection of a Y-specific and X-specific DNA sequence. Int J Leg Med 1993;105:213–6.
5. Oldroyd NJ, Urquhart AJ, Kimpton CP, Millican ES, Watson SK, Downes T, et al. A highly discriminating octoplex short tandem repeat PCR system suitable for human individual identification. Electrophoresis 1995;16:334–7.
6. Mills KA, Even D, Murray JC. Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). Hum Mol Genet 1992;1:779.
7. Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J, et al. A microsatellite genetic linkage map of human chromosome 18. Genomics 1993;15:48–56.
8. Sharma V, Litt M. Tetranucleotide repeat polymorphism at the D21S11 locus. Hum Mol Genet 1992;1:67.
9. Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Meth Appl 1993;3:13–22.
10. Urquhart A, Kimpton CP, Downes TJ, Gill P. Variation in short tandem repeat sequences—A survey of twelve microsatellite loci for use as forensic identification markers. Int J Leg Med 1994;107:13–20.
11. Eng B, Ainsworth P, Wayne JS. Anomalous migration of PCR products using nondenaturing polyacrylamide gel electrophoresis: The amelogenin sex-typing system. J Forensic Sci 1994; 36:1356–9.
12. Akane A, Seki S, Shiono H, Nakamura H, Hasegawa M, Kagawa M, et al. Sex determination of forensic samples by dual PCR amplification of an X-Y homologous gene. Forensic Sci Int 1992;52:143–8.
13. Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative sex test: Fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 1993;15:636–41.
14. Jones DA. Blood samples: Probability of discrimination. J Forensic Sci Soc 1972;12:355–9.
15. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. Genetics 1974;76:379–90.

Additional information and reprint requests:
Florence Rousselet
Institut de Médecine Légale
11 Rue Humann
67085 Strasbourg Cedex
France