Automated DNA Profiling Employing Multiplex Amplification of Short Tandem Repeat Loci

Colin P. Kimpton, Peter Gill, Abbie Walton, Andy Urquhart, Emma S. Millican, and Maia Adams

Central Research and Support Establishment, Forensic Science Service, Aldermaston, Reading, Berks RG7 4PN, United Kingdom

We have employed automated fluorescence-based technology to detect amplified tri-, tetra-, and pentanucleotide short tandem repeat (STR) loci electrophoresed on denaturing polyacrylamide sequencing gels. The system described incorporates an internal size standard in each sample, allowing the STR–PCR products to be sized automatically with a high degree of precision. By utilizing different fluorescent dye markers for loci that have overlapping allele size ranges, we have developed three multiplex STR systems containing a total of 14 different loci. These multiplex systems were then used to evaluate the usefulness of the 14 loci for the identification of individuals. Allele frequency data were collected from a minimum of 50 individuals from each of three different racial groups: Caucasians, Afro-Caribbeans, and Asians. Of the resulting 42 locus population sets, deviation from Hardy–Weinberg equilibria was detected in only the STR HUMCYAR03-Caucasian data. The probabilities of two unrelated individuals matching by chance (pM) at all 14 loci in the three multiplex reactions was $<1 \times 10^{-14}$. The combination of multiplex STR–PCR and automatic fluorescence-based detection is thus a rapid and powerful technique for individual identification.

STR (or microsatellite) loci consist of simple tandemly repeated sequences of 1–6 bp in length. As with the larger variable number tandem repeat (VNTR or minisatellite) loci, STRs may exhibit a high degree of length polymorphism owing to variation in the number of repeat units displayed. However, unlike VNTRs, which occur predominantly in telomeric regions, STRs appear to be abundant throughout the human genome and occur, on average, every 6–10 kb.[1]

Polymorphic STR sequences have been reported in both genic and extragenic regions of the human genome. Those in genic regions are present not only in intron and flanking sequences, but also within coding regions.[2] STR loci are also commonly associated with the 3′ end of *Alu* repeats, presumably resulting from degeneration of the *Alu* poly(A) tail sequence.[1,3]

Because of their abundance, polymorphic nature, and amenability to amplification by PCR, STRs are ideal markers for genomic mapping and genetic linkage analysis. Numerous STR loci have been reported for use in predictive diagnosis of many genetically inherited diseases, including myotonic dystrophy,[4] cystic fibrosis,[5] Duchenne/Becker muscular dystrophy,[6] and Huntington's disease.[7] Genetic linkage maps of the human genome based entirely on dinucleotide STR loci have also been generated.[8] Furthermore, it has recently become apparent that enlargement of polymorphic trinucleotide repeat loci may be the primary cause of a number of genetic diseases such as myotonic dystrophy,[9] X-linked spinal and bulbar muscular atrophy,[10] and fragile X syndrome.[11]

In addition to their suitability for mapping and linkage analysis, STRs provide a source of highly informative loci for use in the identification of individuals. DNA profiling based on PCR amplification of STRs has the advantage of being more sensitive than conventional techniques. Furthermore, because of their small allele sizes (generally < 300 bp), STR systems are more likely to be successful on old or poorly stored specimens that contain only degraded DNA.[12–14] Also, the ability to resolve PCR products differing in size by just 1 base on polyacrylamide gels allows precise allele designation, thus eliminating the need for the continuous allele distribution models currently employed with VNTR systems.[15,16]

Analysis of dinucleotide STRs has revealed enzyme slippage during amplification, resulting in artifactual "stutter" bands. This makes unambiguous allele designation difficult. However, tri- and tetrameric repeats, which have a wider allele spacing, appear to be significantly less prone to slippage and are therefore more suitable for individual identification.[2,17] In addition, the ability to amplify multiple loci in a single "multiplex" reaction,[2,5,18] coupled with the direct detection of amplified products on polyacrylamide gels, makes STR DNA profiling amenable to automation.

The use of automated fluorescence-based technology for the detection of both VNTR and dinucleotide STR loci has been described recently.[18–20] Here, we report the use of similar technology in the evaluation of 14 of the more polymorphic 3- to 5- bp STRs currently available for individual identification, and document an automated fluorescence-based multiplex DNA profiling system.

3:13–22©1993 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/93 $5.00

This system incorporates automatic sizing of PCR products and eliminates differences in electrophoretic mobility among gel lanes by the inclusion of an internal sizing standard with every sample.

MATERIAL AND METHODS

Source of DNA

DNA was prepared from whole blood as described previously.[15] Blood samples were obtained from unrelated Caucasians, Afro-Caribbeans, and Asians residing within the United Kingdom.

Locus-specific Amplification Conditions

Repeat unit and primer sequences for the STR loci under study are given in Table 1. All oligonucleotide primers used were synthesized commercially (Oswel DNA Services). Selected primers were labeled with one of the fluorescent dye markers FAM (5-carboxyfluorescein), JOE (2′,7′-dimethoxy-4′,5′-dichloro-6-carboxyfluorescein), or TAMRA (*N,N,N′,N′*-tetramethyl-6-carboxyrhodamine) [Applied Biosystems Incorporated (ABI)] coupled with an aminohexyl linker (Oswel DNA Services). PCR amplification was performed using 10 ng of genomic DNA in a 50-μl reaction volume. Each locus, with the exception of HUMAPOAI1, was amplified initially in an individual reaction consisting of 1× PARR buffer (Cambio Laboratories), 1.25 units of *Taq* polymerase, 200 μM dNTPs, and 0.25 μM of each relevant primer. HUMAPOAI1 reactions contained 0.065 μM of each primer.

PCR cycling conditions for each locus were as follows: (HUMCD4) 27 cycles at 94°C for 45 sec, 60°C for 30 sec, and 72°C for 30 sec; (HUMDHFR, HUMCYARO3, HUMTH01 and HUMPLA2A) 28 cycles at 94°C for 45 sec, 60°C for 30 sec, 72°C for 30 sec; (HUMAPOAI1 and HUMF13A1) 27 cycles at 94°C for 45 sec, 55°C for 30 sec, and 72°C for 30 sec; (HUMVWA31/A) 28 cycles at 94°C for 45 sec, 50°C for 30 sec, and 72°C for 30 sec; (HUMFES/FPS) 28 cycles at 94°C for 45 sec, 55°C for 30 sec, and 72°C for 30 sec; (HUMGABARB1 and HUMACTBP2) 28 cycles at 94°C for 60 sec, 60°C for 60 sec, and 72°C for 60 sec; (D21S11) 30 cycles at 94°C for 60 sec, 60°C for 60 sec, and 72°C for 60 sec; (HUMFIIDA and HUMFABP) 26 cycles at 94°C for 45 sec, 55°C for 30 sec, and 72°C for 30 sec.

Multiplex PCR

Multiplex 1: HUMVWA31, HUMTH01, HUMF13A1, HUMFES/FPS Reaction components: 1× PARR buffer, 1.25 units of *Taq* polymerase, 200 μM dNTPs, 0.25 μM VWA/1 and 2, 0.10 μM TH01/1 and 2, 0.10 μM F13/1 and 2, and 0.125 μM FES/1 and 2. PCR cycling conditions: 28 cycles at 94°C for 45 sec, 54°C for 30 sec, and 72°C for 30 sec.

Multiplex 2: HUMCD4, HUMDHFR, HUMCYARO3, HUMAPOAII, HUMPLA2A, HUMIIDA, HUMFABP Reaction components: 1× PARR buffer, 1.25 units of *Taq* polymerase, 200 μM dNTPs, 0.25 μM

TABLE 1 STR Loci Under Study and Primer Sequences Employed

Locus	Accession number	Chromosomal location	Primers[a]	Reference
HUMVWA31/A	M25858	12p12-pter	VWA/1 CCCTAGTGGATGATAAGAATAATC TAMRA VWA/2 GGACAGATGATAAATACATAGGATGGATGG	21
HUMTH01	D00269	11p15-15.5	TH01/1 GTGGGCTGAAAAGCTCCCGATTAT FAM TH01/2 GTGATTCCCATTGGCCTGTTCCTC	2,22
HUMF13A1	M21986	6p24-25	F13A1/1 ATGCCATGCAGATTAGAAA JOE F13A1/2 GAGGTTGCACTCCAGCCTTT	23
HUMFES/FPS	X06292	15q25-qter	FES/1 GGGATTTCCCTATGGATTGG FAM FES/2 GCGAAAGAATGAGACTACAT	24
HUMCD4	—	12p	CD4/1 TTGGAGTCGCAAGCTGAACTAGC FAM CD4/2 GCCTGAGTGACAGAGTGAGAACC	25
HUMPLA2A	M22970	12	PLA2A/1 CCCACTAGGTTGTAAGCTCCATGA JOE PLA2A/2 TACTATGTGCCAGGCTCTGTCCTA	26
HUMDHFRP2	J00145	6	DHFR/1 ATTGTAAGACTTTTGGAGCCATTT JOE DHFR/2 TTCAGGGAGAATGAGATGGGC	27
HUMCYARO3	M30795	15q21.1	P450/1 CTCTGGAAAACAACTCGACCCTTC FAM P450/2 TGGGTGATAGAGTCAGAGCCTGTC	28
HUMF11DA	M36960	6	FIIDA/1 GCCTATTCAGAACACCAATA FAM FIIDA/2 TGGGACGTTGACTGCTGAAC	29
HUMFABP	M18079	4q28-31	FABP/1 GTAGTATCAGTTTCATAGGGTCACC FAM FABP/2 TTACGCGTCTCGGACAGTATTCAGTTCGTTTC	2,30
HUMGABARB1	—	4p12-13	GABA/1 CTAGAAAGCTAGCAAGGTGGAT FAM GABA/2 GCTCATTAAACACTGTGTTCCT	31
HUMD21S11	M84567	21	D21/1 ATATGTGAGTCAATTCCCCAAG JOE D21/2 TGTATTAGTCAATGTTCTCCAG	32
HUMACTBP2	V00481	5/6	ACTBP2/1 AATCTGGGCGCACAAGAGTGA FAM ACBTP2/2 ACATCTCCCCTACCGCTATA	33,34
HUMAPOAI1	J00048	11q23-qter	APOAII/1 GGAGCAGTGCTAGGGCCGCGCCGT FAM APOAII/2 GTGACAGAGGGAGACTCCATTAAA	35,36

[a]Fluorescent dye labels (ABI): FAM (blue), JOE (green), TAMRA (yellow).

CD4/1 and 2, 0.50 μM DHFR/1 and 2, 0.125 μM P450/1 and 2, 0.125 μM APOAII 1 and 2, 0.375 μM PLA2A/1 and 2, 0.50 μM IIDA/1 and 2, and 0.065 μM FABP/1 and 2. PCR cycling conditions: 30 cycles at 94°C for 45 sec, 60°C for 30 sec, and 72°C for 30 sec.

Multiplex 3: HUMGABA, D21S11, HUMACTBP2 Reaction components: 1× PARR buffer, 1.25 units of *Taq* polymerase, 200 μM dNTPs, 0.12 μM GABA/1 and 2, 0.5 μM D21S11/1 and 2, and 0.125 μM SE33/1 and 2. PCR cycling conditions: 28 cycles at 94°C for 60 sec, 60°C for 60 sec, and 72°C for 60 sec.

Detection System

A 1- to 4-μl aliquot of each amplification reaction was combined with 6 fmoles of internal lane standard GS2500 (ABI), consisting of *Pst*I-restricted bacteriophage λ DNA labeled with the dye ROX. PCR product and internal standard samples were heat denatured before being loaded onto a standard 6% polyacrylamide denaturing sequencing gel. Gels were electrophoresed for 8 hr at constant power (30 W) on an Applied Biosystems automated DNA sequencer model 370. Fragment sizes were determined automatically using GENESCAN 672 software (ABI) employing the method of second-order regression to establish a curve of best fit for the internal standard in each lane.

Statistical Calculations

The precision of band size estimation by the GENESCAN 672 software was determined for each locus using the formula

$$\% \text{ Precision} = \left[1 - \frac{\text{SD}}{\text{Exp}}\right] \times 100$$

with SD being the standard deviation of observed band sizes for a given allele, and Exp being the theoretical allele size.

Hardy–Weinberg equilibria were tested using the log Likelihood G statistic.[37] Polymorphic information content (PIC) was calculated using the formula of Botstein et al.[38] Discriminating power and matching probabilities (pM) were calculated by the method of Jones.[39] The sample gene diversity (frequency of heterozygotes expected under Hardy–Weinberg equilibrium) was calculated as

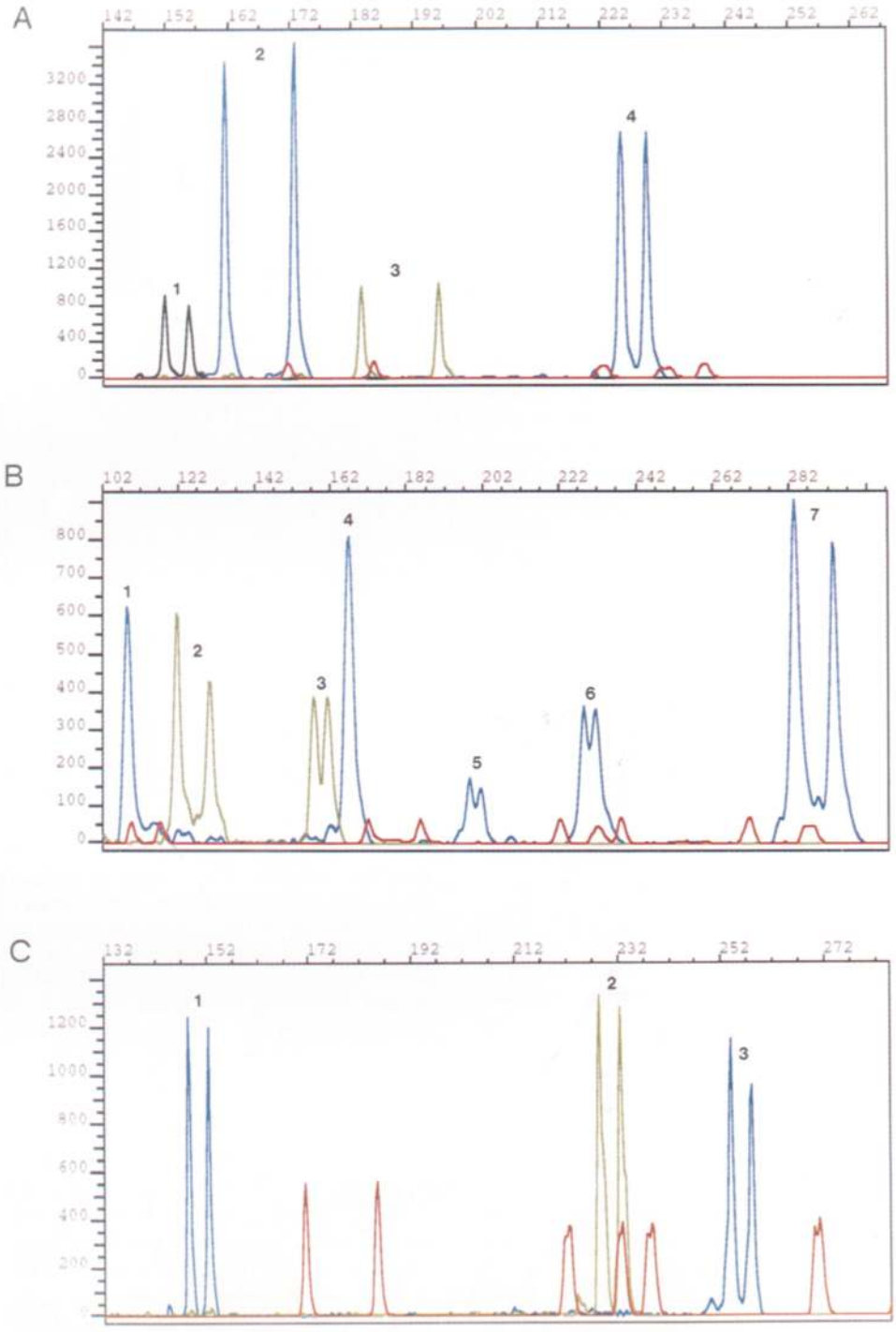

FIGURE 1 Electrophoretograms of individual lanes from a 6% denaturing polyacrylamide gel. Peaks represent fluorescent intensities of dye-labeled DNA products. Peak height is measured against an arbitrary scale displayed on the y axis. The size of the products, in bases, is shown along the x axis. (*A*) Multiplex 1: Locus 1 HUMVWA (black), locus 2 HUMTHO1 (blue), locus 3 HUMF13A1 (green), locus 4 HUMFES/FPS (blue). GS2500 internal size standard (red). (*B*) Multiplex 2: Locus 1 HUMCD4 (blue), locus 2 HUMFIIDA (blue), locus 3 HUMPLA2A (green), locus 4 HUMDHFRP2 (green), locus 5 HUMCYARO3 (blue), locus 6 HUMFABP (blue), locus 7 HUMAPOAI1 (blue). GS2500 internal size standard (red). (*C*) Multiplex 3: Locus 1 HUMGABARB1 (blue), locus 2 D21S11 (green), locus 3 HUMACTBP2 (blue). GS2500 internal size standard (red).

$$d = 1 - \sum_i pi^2$$

with pi being the allele frequency.

RESULTS

Fourteen 3- to 5-bp STR loci were selected for evaluation based on their predicted discrimination power, as indicated by data published previously (for references, see Table 1).

STR amplification products were tagged by the attachment of a fluorescent dye molecule to one of each pair of locus-specific primers. Amplified products were then detected by laser scanning during electrophoresis on denaturing polyacrylamide gels. Band sizes were generated automatically by comparison with a standard sizing ladder included in every sample prior to electrophoresis.

PCR component concentrations and cycling parameters were optimized for each locus individually (see Materials and Methods). STRs suitable for coamplification (multiplexing) were then selected on the basis of similar optimal reaction conditions and compatible allele size ranges. Three multiplex systems were constructed, containing a total of 14 loci. Examples of electrophoretograms displaying the PCR products of these multiplex reactions are shown in Figure 1. STR loci with overlapping allele size ranges were differentiated by use of different fluorescent dye labels. Comparable signal intensities of amplified products within a multiplex system were obtained by adjustment of individual primer concentrations (see Materials and Methods).

Microsatellite loci were found to coamplify with relative ease, and all multiplex systems employed identical buffer, dNTP, and enzyme concentrations. Cycling times and temperatures were also similar, with only the annealing temperature requiring adjustment for each specific multiplex system.

Computer-generated band sizes for 12 of the 14 loci fell into discrete groups. This is demonstrated in Figure 2A by the HUMTH01 size calling data, which is representative of the 12 loci. The difference between the mean band sizes of each group corresponded to the theoretical repeat unit size for 11 of the 12 loci. The maximum band size range for any given group was 1.2 bp [corresponding to ± 0.6 bp either side of the mean (Table 2).] As the minimum repeat unit size was 3 bp, unambiguous allele designation was possible for every sample for the 11 loci mentioned above. The microsatellite D21S11 displayed a number of allele bands differing by only 2 bp (theoretical repeat unit size was 4 bp). However, because the maximum band size range for this locus was only 0.7 bp (0.35 bp either side of the mean), these half repeat unit alleles could be identified precisely. Therefore, it was also possible to readily assign allele designations for D21S11.

The remaining two loci (HUMAPOAI1 and HUMACTBP2) displayed a number of allele bands differing by 1 bp and 2 bp, respectively. Although such differences were readily resolvable on polyacrylamide gels, the consistency of automatic sizing between gels was not sufficient to allow precise allele designation using these loci (demonstrated for HUMACTBP2 in Fig. 2B). For this reason, allele assignment for HUMAPOAI1 and HUMACTBP2 was undertaken by the direct comparison of computer-generated band sizes to those of an allelic ladder control run on the same gel.

Allele bands for all loci were sized with a high degree of precision when

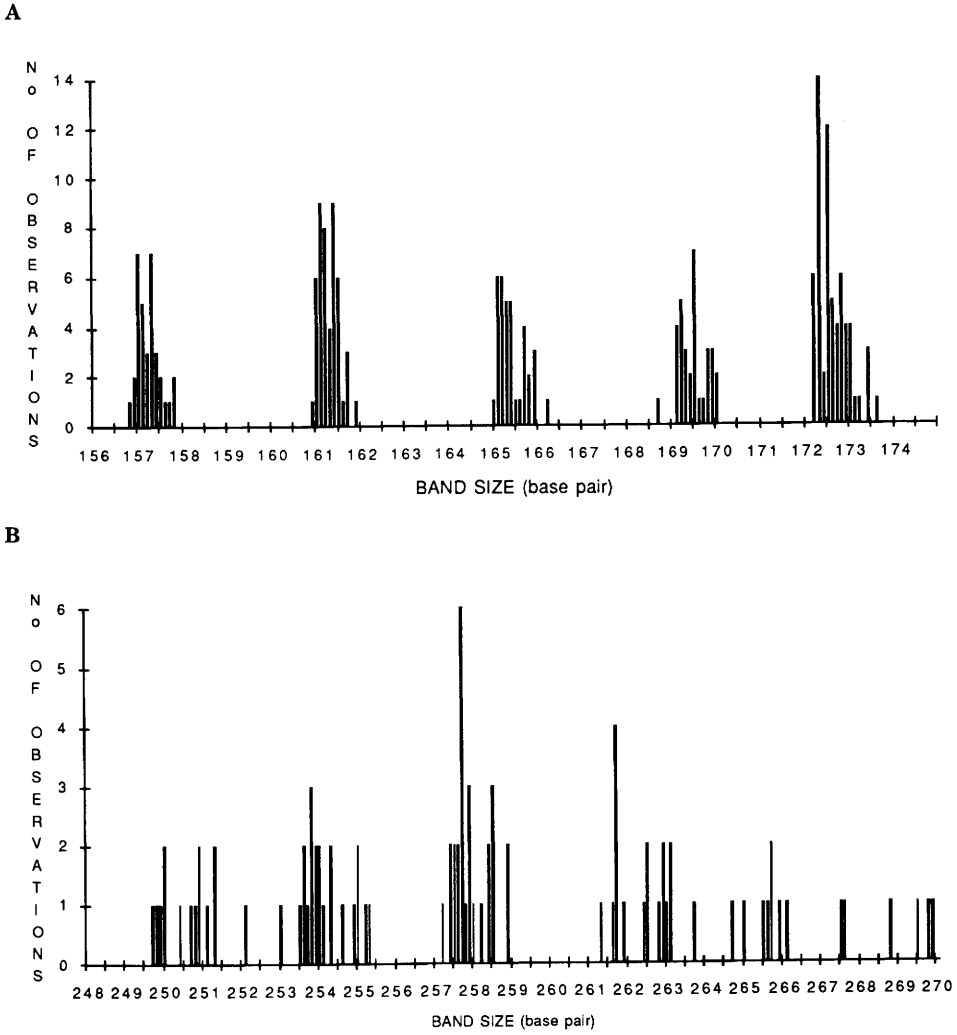

FIGURE 2 Frequency distributions of automatically generated PCR product sizes. Amplification products were combined with 6 fmoles of an internal lane standard before being loaded onto 6% polyacrylamide denaturing gels. Fragment sizes were determined automatically using GENESCAN 672 software (ABI) employing the method of second-order regression to establish a curve of best fit for the internal standard in each lane. (*A*) HUMTHO1 sizing data for 212 allele bands from six gels. (*B*) HUMACTBP2 sizing data for 92 allele bands between the range of 248–270 bases in length, from five gels.

TABLE 2 Precision of Computer-generated Band Sizing

Locus	Within gels			Across gels		
	precision (%)	s.d.[a] (bp)	range[b] (bp)	precision (%)	s.d.[a] (bp)	range[b] (bp)
HUMCD4	99.92	(0.10)	0–0.6	99.78	(0.26)	0.5–1.2
HUMDHFR	99.94	(0.10)	0.1–0.6	99.89	(0.18)	0.7–1.0
HUMCYARO3	99.95	(0.09)	0.1–0.5	99.87	(0.24)	0.6–1.1
HUMVWA31/A	99.93	(0.10)	0–0.7	99.82	(0.27)	0.6–1.2
HUMTH01	99.95	(0.08)	0–0.5	99.93	(0.12)	0.4–0.6
HUMF13A1	99.95	(0.10)	0–0.7	99.88	(0.25)	0.3–1.2
HUMFES/FPS	99.95	(0.11)	0–0.7	99.92	(0.19)	0.1–0.9
HUMPLA2A	99.95	(0.07)	0.1–0.4	99.90	(0.12)	0.4–0.8
HUMF11DA	99.95	(0.10)	0.1–0.5	99.87	(0.24)	0.4–1.1
HUMFABP	99.95	(0.12)	0.2–0.7	99.87	(0.31)	0.2–1.2
HUMGABAR1	99.95	(0.08)	0–0.6	99.90	(0.14)	0.2–0.8
HUMD21S11	99.96	(0.09)	0–0.6	99.94	(0.13)	0.1–0.7
HUMACTBP2		ND		99.79	(0.53)	0.7–2.1
HUMAPOA11		ND		99.84	(0.44)	0.1–1.4

[a]Standard deviation. (ND) Not done.
[b]Range of computer generated sizes for a given allele.

compared within gels (Table 2). However, this level of precision was reduced when sizes were compared among gels, although the reduction was only significant with the HUMAPOAI1 and HUMACTBP2 loci. With these STRs the migration of PCR product bands appeared to be affected by gel to gel variation to a greater extent than the internal size standard bands. This is probably because of the substantially different sequence compositions of the PCR products and the size standard; both HUMAPOAI1 and HUMACTBP2 contain long AT-rich repeat regions.

Allele frequencies for each STR locus under investigation were determined from a minimum of 50 random individuals for each of three different populations: Caucasians, Afro-Caribbeans, and Asians. Allele frequency histograms for all 14 loci are shown in Figure 3. Symmetrical and skew unimodal, bimodal and more complex distributions were observed among the 14 loci. Differences in allele frequencies among population groups were also seen. This was most pronounced for HUMFIIDA, with the smaller alleles (1–8) being significantly more common in the Afro-Caribbean population. The overall incidence of these alleles in Afro-Caribbeans was 58%, in contrast with 5% and 10% in Caucasians and Asians, respectively. Allele frequency differences among populations were compared for each microsatellite using a Chi-square test. Interpopulation differences were significant ($P < 0.05$) for all loci, with the exception of HUMFES/FPS ($P = 0.1$) and HUMGABARB1 ($P = 0.38$).

The data sets were tested for Hardy–Weinberg equilibria using a log likelihood-G test.[37] In total, 42 locus population comparisons were carried out. Deviation from Hardy–Weinberg equilibria was only detected for the HUMCYARO3 Caucasian data ($P < 0.05$).

The number of alleles observed, gene diversity, and PIC for each locus and population are summarized in Table 3. The STR loci appear to fall into two main groups. Those in the first group, which contains all loci except HUMACTBP2, HUMAPOAI1, and D21S11, display a small number of well-separated alleles (<12). These alleles differ in size by the expected length of the repeat unit in virtually every case, with only two exceptions. The exceptions are the smallest allele for HUMF13A1 (4-bp repeat unit), which is only 2 bp less than the second allele; and allele 5 for HUMTHO1 (containing 10 repeat units), which appears to be only 3-bp larger than allele 4. A rare, full-length 10 repeat allele (1 bp larger than the common 10-repeat unit allele) was also observed in this study. Both of these 10-repeat unit alleles were grouped as one in this paper.

The second STR locus group, containing HUMACTBP2 and HUMAPOAI1, is significantly more polymorphic and displays a greater number of alleles (>35). However, these loci appear to be much more complex and possess multiple alleles, which differ by less than the published repeat unit size. The D21S11 locus falls between the two groups.

The pMs of individual loci (Table 4) was calculated from the allele frequency data using the formula of Jones.[38]. Values ranged from 0.25 for HUMFABP in Asians to 0.02 for both HUMACTBP2 in Caucasians and HUMAPOAI1 in Afro-Caribbeans and Asians. Differences in discriminating power between populations occurred to varying degrees for individual loci. However, in most cases the loci were more discriminating for Afro-Caribbeans. The latter population also tended to display a greater number of alleles for most loci, indicating that the Afro-Caribbean population was more heterogeneous than the Caucasian or Asian groups. This has also been reported for VNTRs.[40] The average combined pMs of the multiplex reactions, calculated by the product rule, were: 1.3×10^{-4} for multiplex 1 (4 loci); 5×10^{-7} for multiplex 2 (7 loci); and 1.6×10^{-4} for multiplex 3 (3 loci). The combined figure for the three multiplex systems is thus $>1 \times 10^{-14}$.

Artifactual stutter bands caused by enzyme slippage during amplification were detected with a number of STR loci. The degree of slippage appeared to be locus dependent and remained relatively constant, ranging from negligible (e.g., HUMTH01) to a maximum of 12% of the signal intensity of the adjacent allele as measured by peak area (e.g., HUMPLA2A).

Variation in signal intensity of different allele bands in heterozygote samples was also observed for some loci. Again, this variation appeared to be locus dependent, and in most instances the larger of the two alleles gave a less intense signal. This was most apparent when allele size differences were pronounced (>50 bp) and was probably the result of preferential amplification of smaller sequences during PCR. Three loci (HUMCD4, HUMCYARO3, and HUMFES/FPS) displayed signal variation for either of the two alleles. The reason for this is unclear. Preliminary studies of mixed samples using loci that showed negligible signal variation indicated that signal intensity can be a relatively accurate measure of template DNA ratios (data not shown).

The addition of an extra base by the *Taq* polymerase enzyme at the end of elongation can result in a PCR product 1

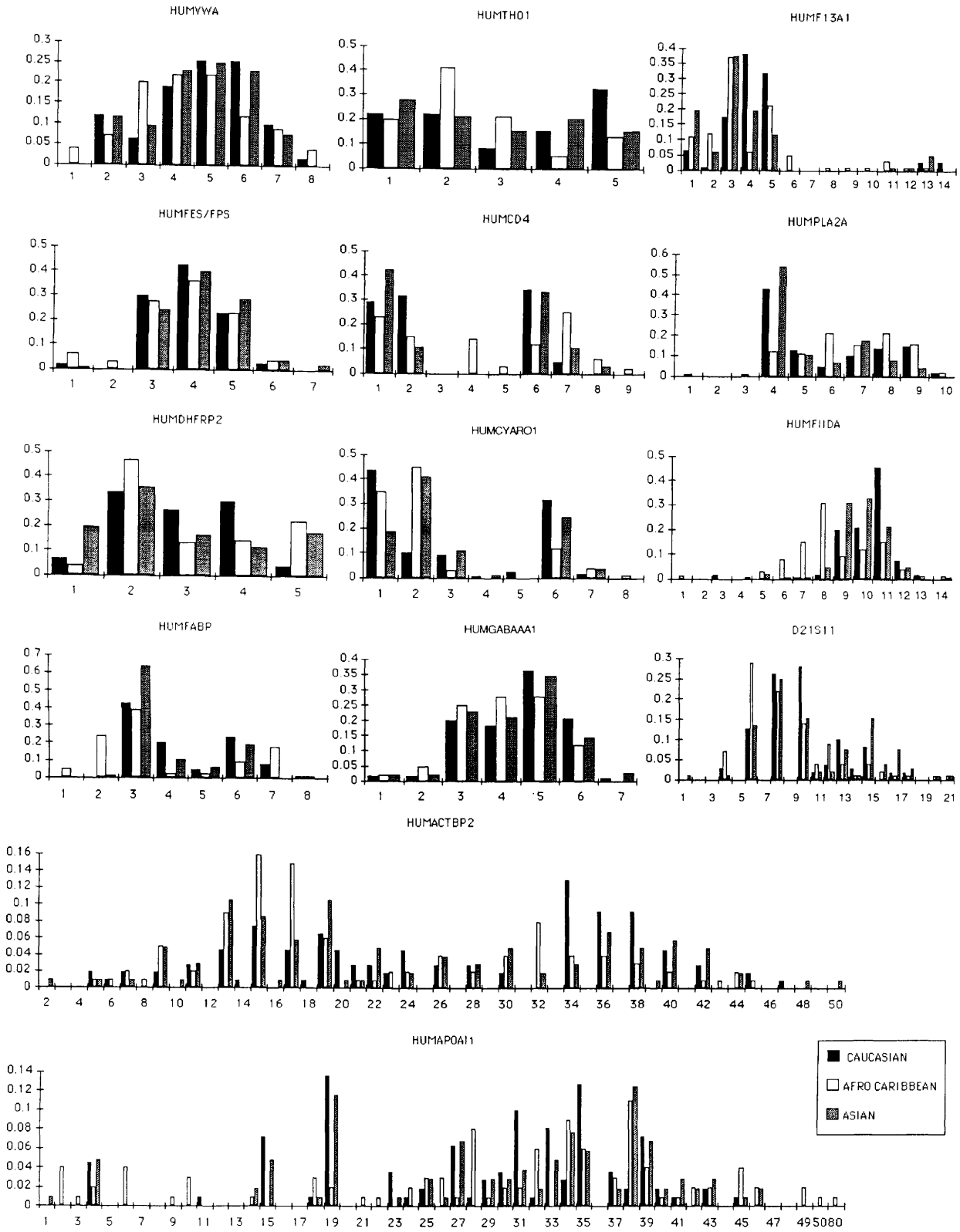

FIGURE 3 Allele frequency histograms of the 14 STR loci studied. The data were generated from a minimum of 50 individuals from each of three populations: Caucasians, Afro-Caribbeans, and Asians. Allele designation number is represented on the x axis. Allele frequency is shown on the y axis.

base larger than its predicted length.[41] This addition does not occur with every product copy, and double peaks that differ by 1 base may be seen for each allele. The ratio of true length to true length plus 1 base product again appeared to be largely locus dependent. Examination of three different strategies for elimination of double bands, due to extra base addi-

TABLE 3 Number of Alleles, Gene Diversity, and PIC of the STR Loci Studied

Locus	Caucasian (N = 55)[a]			Afro-Caribbean (N = 50)[a]			Asian (N = 55)[a]			Total K
	K	gene D	PIC	K	gene D	PIC	K	gene D	PIC	
HUMVWA31/A	7	0.81	0.79	8	0.83	0.82	6	0.80	0.79	8
HUMTH01	5	0.77	0.76	5	0.73	0.70	5	0.78	0.78	5
HUMF13A1	7	0.72	0.68	12	0.78	0.78	8	0.77	0.75	13
HUMFES/FPS	5	0.67	0.62	6	0.73	0.70	6	0.69	0.65	7
HUMCD4	4	0.69	0.68	8	0.82	0.82	5	0.68	0.68	8
HUMPLA2A	7	0.75	0.74	9	0.84	0.83	6	0.66	0.65	9
HUMDHFR	5	0.72	0.69	5	0.69	0.68	5	0.77	0.75	5
HUMCYARO3	7	0.69	0.69	6	0.66	0.61	6	0.72	0.71	8
HUMFIIDA	8	0.70	0.68	11	0.83	0.82	10	0.75	0.72	13
HUMFABP	6	0.71	0.70	8	0.75	0.73	5	0.55	0.54	8
HUMGABARB1	7	0.75	0.73	6	0.76	0.74	7	0.76	0.74	7
HUMD21S11	11	0.81	0.81	14	0.83	0.83	15	0.86	0.86	15
HUMACTBP2	26	0.94	0.94	26	0.92	0.92	28	0.94	0.94	36
HUMAPOAI1	24	0.92	0.92	34	0.95	0.95	26	0.94	0.94	39

[a](N) Number of individuals studied for each locus. (K) Number of alleles observed. (Gene D) Gene diversity (expected heterozygosity).

tion, was carried out. The first strategy involved incorporation of an *Mlu*I restriction site into the non-dye-labeled primer sequence. This allowed cleavage of a small sequence, including the extra base if it had been added, from one end of the PCR product (Fig. 4). This procedure was successful, although small bands representing a proportion of uncleaved product were still detectable. A simpler method to remove the extra base involved the addition of 2 units of T4 DNA polymerase to the PCR product followed by incubation at 37°C for an additional 30 min. The extra base was efficiently removed by the 3′ to 5′ single-stranded exonuclease activity of the T4 DNA polymerase (Fig. 4). The final strategy examined was adjustment of the amplification conditions to favor addition of the extra base, thus eliminating the formation of double peaks. This was achieved by including a final 10-min. incubation at 72°C and increasing the starting concentration of dATP. Although this did not completely eliminate the true length band for every locus, it was sufficient to allow unambiguous allele designation.

TABLE 4 Matching Probability (pM) of STR Loci

Locus	Caucasian	Afro-Caribbean	Asian
Multiplex 1			
HUMVWA31/A	0.07	0.06	0.08
HUMTH01	0.09	0.15	0.11
HUMF13A1	0.13	0.09	0.11
HUMFES/FPS	0.19	0.15	0.14
Combined pM	1.5×10^{-4}	1.2×10^{-4}	1.3×10^{-4}
Multiplex 2			
HUMCD4	0.16	0.07	0.06
HUMPLA2A	0.10	0.07	0.15
HUMDHFR	0.14	0.15	0.10
HUMCYARO3	0.16	0.17	0.14
HUMFIIDA	0.13	0.16	0.13
HUMFABP	0.16	0.11	0.25
HUMAPOAI1	0.03	0.02	0.02
Combined pM	2.2×10^{-7}	1.6×10^{-8}	2.5×10^{-7}
Multiplex 3			
HUMGABARB1	0.10	0.10	0.10
HUMD21S11	0.07	0.07	0.05
HUMACTBP2	0.02	0.08	0.03
Combined pM	1.4×10^{-4}	2.1×10^{-5}	1.5×10^{-4}
Σ Combined pM	4.6×10^{-15}	4.0×10^{-16}	4.9×10^{-15}

DISCUSSION

We have developed multiplex amplification systems that were used in conjunction with automated fluorescence-based detection to evaluate the suitability of 14 highly polymorphic tri-, tetra-, and pentanucleotide STR loci for identification of individuals.

The STR loci studied were shown to coamplify readily under standard PCR conditions. Efficient amplification of all loci in multiplex systems was achieved by the adjustment of annealing temperature and individual primer concentration. In contrast, VNTR loci often require diverse reaction conditions for successful amplification and are difficult to coamplify.[19,42] Furthermore, allelic dropout caused by preferential amplification of smaller sequences has been reported with some amplified VNTR loci.[42] Allelic dropout does not appear to occur with STR loci, because of their narrow allele size range.

The narrow allele size range of STRs also increases the potential number of loci that can be differentiated in one gel lane. This may be further facilitated by the use of different fluorescent dye markers for loci with overlapping allele size ranges. The availability of four distinguishable fluorescent dyes also allows the incorporation of an internal size standard in every gel lane, thereby eliminating interpretation errors caused by differences in electrophoretic mobility across a gel. Moreover, by employing the GENESCAN 672 computer software, it was possible to size PCR products automatically against the internal ladder standard. The data were then stored in a spreadsheet format. The use of a universal internal size standard eliminates the requirement to develop allelic ladders for each separate locus and, thus, allows every lane of a gel to be used for experimental samples.

The precision achieved by the automatic sizing software was sufficient to allow unambiguous allele designation for 12 of the 14 loci studied. However, variability between gels did not allow reliable allele designation for two STRs (HUMAPOAI1 and HUMACTBP2), which

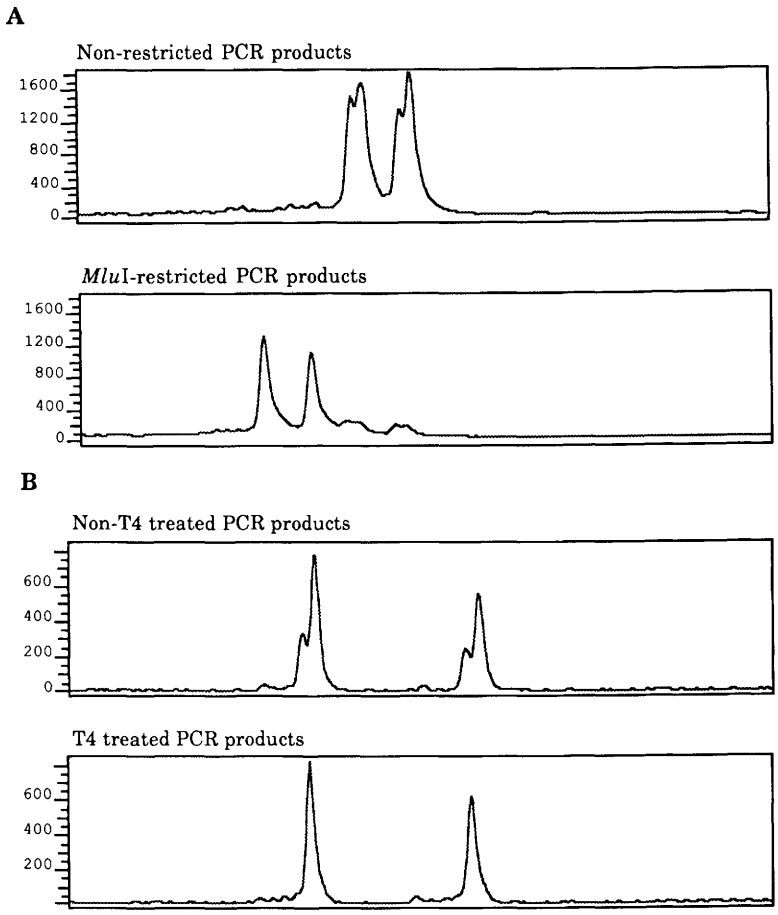

FIGURE 4 Incomplete addition of an extra base by the *Taq* polymerase enzyme at the end of elongation results in double peaks differing by 1 base for each allele. (*A*) Elimination of double peaks by incorporation of an *Mlu*I restriction site in the non-dye-labeled primer coupled with *Mlu*I restriction of products following amplification. (*B*) Removal of the extra base by incubation of PCR products with T4 DNA polymerase.

contain long AT-rich repeat regions and display alleles differing by only 1 and 2 bp. With both of these loci it was necessary to run an allelic ladder on each gel. The anomalous migration rates of amplification products relative to an internal size standard have also been reported for AT-rich VNTRs.[19] This is probably caused by differences in DNA-gel interactions, which are affected by variation in gel composition, and by a DNA-bending phenomenon to which AT-rich sequences are more susceptible.[43] This problem could be overcome by use of an AT-rich internal size standard for AT-rich loci, as opposed to the use of specific allelic ladder markers.

Examination of the sequence of HUMAPOAI1 and HUMACTBP2 reveals that the repeat regions of these loci consist of degenerative repeat sequences of varying size.[44,45] Moreover, length variation appears to occur at a number of sites within the repeat regions of these loci,[46] thus explaining the presence of alleles differing by 1 and 2 bp. Prior to routine use of these loci by forensic laboratories for the identification of individuals, it must be confirmed that the detection and sizing protocols used allow accurate, reliable, and unambiguous allele designation.

Sequence analysis of D21S11 has revealed a common 2-bp mutation within the repeat region of this locus (A. Urquhart, pers. comm.), which would account for the 2-bp allele differences observed for D21S11.

The reason for the different STR frequency distributions observed here (unimodal, bimodal, and complex) is unclear. Such distributions have been reported previously for STR loci, and it was suggested that they may represent the evolutionary history of the alleles for different populations.[17] In addition, structural limitations on allele sizes are also likely to play a major role in the generation and frequency of specific alleles.

Interpopulation differences in allele frequencies were significant for 12 of the 14 loci examined in this study, indicating the potential usefulness of STR loci as race markers. A number of loci appeared to display race specific alleles, although these alleles occurred at a relatively low frequency. It would be necessary to type substantially greater numbers of individuals before such alleles could be confirmed as being completely race-specific. It may be possible to group potential race-informative loci in a single multiplex reaction allowing a probability of racial origin to be calculated from the profile of an individual of unknown origin.

Although the number of chromosomes studied for each locus and population was relatively small in this study, preliminary tests indicated that with the exception of HUMCYARO3 in Caucasians, all of the loci were within Hardy–Weinberg equilibria. The apparent deviation from the Hardy–Weinberg equilibrium for the HUMCYARO3 data is currently under more detailed investigation but may be simply because of sampling error. The overall pM of the three multiplex systems developed was $<1 \times 10^{-14}$, thus highlighting the power of these systems for individual identification. However, before a discriminatory system is accepted for routine forensic use, it must be proved to be both robust and reliable. The high incidence of artifactual stutter bands observed with dinucleotide STR loci make them unsuitable for forensic applications. However, the significantly reduced level of stuttering for 3- to 5-bp repeat loci observed both in this and previous studies,[2] suggests that these STRs are more amenable to forensic investigations. This is reinforced by preliminary studies carried out in our laboratories on forensic material (data not shown).

Fluorescence signals are known to be linear over a much greater range of intensities than conventional autoradiography.[20] Therefore, fluorescence-based detection systems are likely to be more useful than autoradiography for direct quantitation of PCR products. This has a direct forensic significance when it is necessary to type samples of mixed ori-

gin. Preliminary studies on mixed samples suggest that comparison of fluorescent STR–PCR band intensities may give a relatively accurate quantitation of the proportions of material from each individual in a mixed sample. Currently, we are undertaking detailed studies of mixed forensic specimens which, in turn, should assist in the interpretation of profiles from complicated forensic samples.

In conclusion, tri-, tetra-, and pentanucleotide STR loci appear to be well suited for routine identification of individuals. These loci readily coamplify, thus facilitating the development of highly discriminating multiplex systems. Employment of automated fluorescent detection systems increases the potential for development of multiplex reactions. This can be accomplished by the use of different fluorescent dyes for loci with overlapping allele sizes and the incorporation of an internal size standard in each sample. PCR products can then be automatically sized with a high degree of precision and the data transferred into computer data bases or manipulated further. Thus, the combination of multiplex STR–PCR and automated fluorescent detection results in a rapid and powerful DNA profiling technique with considerable potential in a variety of areas including forensics.

REFERENCES

1. Beckmann, J.S. and J.L. Weber. 1992. Survey of human and rat microsatellites. *Genomics* **12:** 627–631.
2. Edwards, A., A. Civitello, H.A. Hammond, and C.T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49:** 746–756.
3. Economou, E.P., A.W. Bergen, A.C. Warren, and S.E. Antonarakis. 1990. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci.* **87:** 2951–2954.
4. Mully, J.C., A.K. Gedeon, S.J. White, E.A. Haan, and R.I. Richards. 1991. Predictive diagnosis of myotonic dystrophy with flanking microsatellite markers. *J. Med. Genet.* **28:** 448–452.
5. Morral, N. and X. Estivill. 1992. Multiplex PCR amplification of three microsatellites within the CFTR gene. *Genomics* **13:** 1362–1364.
6. Oudet, C., R. Heilig, A. Hanauer, and J.-L. Mandel. 1991. Nonradioactive assay for new microsatellite polymorphisms at the 5′ end of the dystrophin gene, and estimation of intragenic recombination. *Am. J. Hum. Genet.* **49:** 311–319.
7. Weber, B., O. Riess, G. Wolff, S. Andrew, C. Collins, R. Graham, J. Thielman, and M.R. Hayden. 1992. Delineation of a 50 kilobase segment containing the recombination site in a sporadic case of Huntington's disease. *Nature Genet.* **2:** 216–222.
8. Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix, and M. Lathrop. 1992. A second-generation linkage map of the human genome. *Nature* **359:** 794–801.
9. Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy, S. Leblond, J. Earle-Macdonald, P.J. de Jong, B. Wieringa, and R.G. Korneluk. 1992. Myotonic dystrophy mutation: An unstable CTG repeat in the 3′ untranslated region of the gene. *Science* **255:** 1253–1255.
10. La Spada, A.R., E.M. Wilson, D.B. Lubahn, A.E. Harding, and K.H. Fischbeck. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atropy. *Nature* **352:** 77–79.
11. Richards, R.I. and G.R. Sutherland. 1992. Fragile X syndrome: The molecular picture comes into focus. *Trends Genet.* **8:** 249–253.
12. Hagelberg, E., I.C. Gray, and A.J. Jeffreys. 1991. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* **352:** 427–429.
13. Gill, P., C.P. Kimpton, and K.M. Sullivan. 1992. A rapid polymerase chain reaction method for identifying fixed specimens. *Electrophoresis* **13:** 173–175.
14. Jeffreys, A.J., M.J. Allen, E. Hagelberg, and A. Sonnberg. 1992. Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Sci. Int.* **56:** 65–76.
15. Gill, P., K. Sullivan, and D.J. Werrett. 1990. The analysis of hypervariable DNA profiles: Problems associated with the objective determination of the probability of a match. *Hum. Genet.* **85:** 75–79.
16. Evett, I.W. and P. Gill. 1991. A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. *Electrophoresis* **12:** 226–230.
17. Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12:** 241–253.
18. Schwartz, L.S., J. Tarleton, B. Popovich, W.K. Seltzer, and E.P. Hoffman. 1992. Fluorescent multiplex linkage analysis and carrier detection for Duchenne/Becker muscular dystrophy. *Am. J. Hum. Genet.* **51:** 721–729.
19. Sullivan, K.M., S. Pope, P. Gill, and J.M. Robertson. 1992. Automated DNA profiling by fluorescent labeling of PCR products. *PCR Methods Applic.* **2:** 34–40.
20. Ziegle, J.S., Y. Su, K.P. Corcoran, L. Nie, P.E. Mayrand, L.B. Hoff, L.J. McBride, M.N. Kronick, and S.R. Diehl. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14:** 1026–1031.
21. Kimpton, C.P., A. Walton, and P. Gill. 1992. A further tetranucleotide repeat polymorphism in the vWF gene. *Hum. Mol. Genet.* **1:** 287.
22. Polymeropoulos, M.H., H. Xiao, D.S. Rath, and C.R. Merril. 1991. Tetranucleotide repeat polymorphism at the human tyrosine hydrolase gene (TH). *Nucleic Acids Res.* **19:** 3753.
23. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1991. Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). *Nucleic Acids Res.* **19:** 4036.
24. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1991. Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). *Nucleic Acids Res.* **19:** 4018.
25. Edwards, M.C., P.R. Clemens, M. Tristan, A. Pizzuti, and R.A. Gibbs. 1991. Pentanucleotide repeat length polymorphism at the human CD4 locus. *Nucleic Acids Res.* **19:** 4791.
26. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1990. Trinucleotide repeat polymorphism at the human pancreatic phospholipase A-2 gene (PLA2). *Nucleic Acids Res.* **18:** 7468.
27. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1991. Tetranucleotide repeat polymorphism at the human dihydrofolate reductase psi-2 pseudogene (DHFRP2). *Nucleic Acids Res.* **19:** 4792.
28. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1991. Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). *Nucleic Acids Res.* **19:** 195.
29. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1991. Trinucleotide repeat polymorphism at the human transcription factor IID gene. *Nucleic Acids Res.* **19:** 4037.
30. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1990. Trinucleotide repeat polymorphism at the human intestinal fatty acid binding protein gene (FABP2). *Nucleic Acids Res.* **18:** 7198.
31. Dean, M., S. Lucas-Derse, A. Bolos, S.J. O'Brien, E.F. Kirkness, C.M. Fraser, and D. Goldman. 1991. Genetic mapping of the β1 GABA receptor gene to human chromosome 4, using a tetranucleotide repeat polymorphism. *Am. J. Hum. Genet.* **49:** 621–626.
32. Sharma, V. and M. Litt. 1992. Tetranucle-

otide repeat polymorphism at the D21S11 locus. *Hum. Mol. Genet.* **1:** 67.
33. Warne, D., C. Warkins, P. Bodfish, K. Nyberg, and N.K. Spurr. 1991. Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene 2 (ACTBP2) detected using the polymerase chain reaction. *Nucleic Acids Res.* **19:** 6980.
34. Polymeropoulos, M.H., D.S. Rath, H. Xiao, and C.R. Merril. 1992. Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta- Ac-psi-2 (ACTBP2). *Nucleic Acids Res.* **20:** 1432.
35. Zuliani, G. and H. Hobbs. 1990. Tetranucleotide repeat polymorphism in the apolipoprotein C-III gene. *Nucleic Acids Res.* **18:** 4299.
36. Hata, A. and J.-M. Lalouel. 1991. MicroVNTR (CTTT)n at the apo C-III locus. *Nucleic Acids Res.* **19:** 5098.
37. Sokal, R.R. and J.F. Rohlf. 1981. *Biometry*, 2d ed. W.H. Freeman, New York.
38. Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32:** 182–190.
39. Jones, D.A. 1972. Blood samples: Probability of discrimination. *J. Forensic Sci. Soc.* **12:** 355–359.
40. Gill, P., S. Woodroffe, J.E. Lygo, and E.S. Millican. 1991. Population genetics of four hypervariable loci. *Int. J. Leg. Med.* **104:** 221–227.
41. Clark, J.M. 1988. Novel non-templated nucleotide addition reactions catalysed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16:** 9677–9686.
42. Tully, G., K.M. Sullivan, and P. Gill. 1993. Analysis of 6 VNTR loci by multiplex PCR and automated fluorescent detection. (in prep.)
43. Hagerman, P.J. 1990. Sequence directed curvature of DNA. *Annu. Rev. Biochem.* **59:** 755–781.
44. Protter, A.A., B. Levy-Wilson, J. Miller, G. Bencen, T. White, and J.J Seilhamer. 1984. Isolation and sequence analysis of the human apolipoprotein C-III gene and the intragenic region between the Apo AI and Apo CIII genes. *DNA* **3:** 449–456.
45. Moos, M. and D. Gallwitz. 1983. Structure of two beta-actin-related processed genes one of which is located next to a simple repetitive sequence. *EMBO. J.* **2:** 757–761.
46. Urquhart, A., C.P. Kimpton, and P. Gill. 1993. Sequence variation within the tetranucleotide repeat polymorphism at the human beta-actin related pseudogene 2 (ACTBP2). (in prep.).

Received April 23, 1993; accepted in revised form May 4, 1993.

Automated DNA profiling employing multiplex amplification of short tandem repeat loci.

C P Kimpton, P Gill, A Walton, et al.

Genome Res. 1993 3: 13-22