



Automated DNA profiling by fluorescent labeling of PCR products.

K M Sullivan, S Pope, P Gill, et al.

Genome Res. 1992 2: 34-40

Access the most recent version at doi:[10.1101/gr.2.1.34](https://doi.org/10.1101/gr.2.1.34)

References This article cites 21 articles, 3 of which can be accessed free at:
<http://genome.cshlp.org/content/2/1/34.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © Cold Spring Harbor Laboratory Press

Automated DNA Profiling by Fluorescent Labeling of PCR Products

Kevin M. Sullivan, Susan Pope, Peter Gill, and James M. Robertson¹

Central Research and Support Establishment, The Forensic Science Service, Aldermaston, Reading, Berks RG7 4PN, UK;

¹Applied Biosystems, Foster City, California 94404

DNA profiling has been automated by the fluorescent tagging of amplified variable number tandem repeat (VNTR) loci. This was achieved by the use of fluorescently labeled primers in the amplification of 10 ng of genomic DNA, coupled with laser detection of the products during electrophoresis. The PCR products are sized by co-electrophoresing a standard size ladder mixed with every sample, thereby eliminating errors in size estimation caused by lane-to-lane differences in migration rate. This increases the precision of VNTR characterization and enables alleles that differ by a single 15-bp repeat to be resolved. The system is capable of high throughput: Twenty-four samples are electrophoresed and analyzed within 6 hr. Also, because four different dyes are available, three different loci can be simultaneously characterized with the fourth dye used for the internal standard. Approximately 100 unrelated British caucasians were analyzed at the loci D1S80, D17S5, and ApoB. The probabilities of two unrelated individuals matching by chance (pM) at these three loci were determined to be 0.065, 0.040, and 0.069, respectively, with a combined pM of 1.8×10^{-4} .

Variable number tandem repeat (VNTR) loci display considerable variation within human populations and are useful markers for the construction of human linkage maps⁽¹⁾ and in the identification of individuals for forensic purposes.⁽²⁾ VNTRs were originally characterized by restriction enzyme digestion of genomic DNA, followed by agarose gel electrophoresis, Southern blotting, and hybridization with labeled minisatellite probe. PCR amplification of these loci has greatly improved the sensitivity of the technique, enabling as little as 1 ng of genomic DNA to be typed.⁽³⁾ The most informative VNTR loci commonly used in DNA profiling, such as D7S22 and D1S8, have alleles of several kilobases in size.^(4,5) Thus, because of the inverse correlation between the efficiency of amplification versus the length of PCR product, these large amplification products still require Southern blotting and hybridization with an appropriate probe to be detected. Conversely, the amplified alleles from small VNTR loci such as D17S5 and D1S80 can be directly visualized and resolved on ethidium bromide-stained agarose or acrylamide gels, which significantly reduces analysis time.⁽⁶⁻⁹⁾ However, when distinguishing between alleles differing by a single short repeat, such as D1S80 (16 bp) and ApoB (15 bp), it is often essential to run allelic size ladders in adjacent lanes.

We describe here the use of fluorescent tagging of the PCR products coupled with their detection by laser scanning during electrophoresis to increase the precision and to provide automatic characterization of three VNTR loci: D1S80, D17S5, and the hypervariable region 3' to ApoB. Such technology is routinely used in sequence analysis and has been applied recently to other applications, such as quantitative determina-

tion of Duchenne muscular dystrophy status.⁽¹⁰⁻¹⁶⁾ It is ideally suited to VNTR analysis because an internal sizing ladder labeled with a distinguishable dye is included in every lane. This allows the software to size the PCR products automatically based on the DNA migration rate in the individual lanes. Thus, with this system, lane-to-lane electrophoretic differences do not affect the precision of the size calls. Furthermore, since four different dyes are available, it is possible to analyze independently amplification products from three different loci electrophoresed simultaneously in the same lane, thereby providing a highly discriminating test from a single electrophoretic run.

MATERIALS AND METHODS

Amplification Conditions for D17S5, D1S80, and ApoB

For all three loci, ~10 ng of genomic DNA was used in each amplification. Locus D1S80 was amplified in the following reaction mix: 200 μ M each dNTP (BCL), 2.5 units of *Taq* polymerase (CamBio), 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 1% Triton X-100, and 0.1 μ M each primer⁽⁷⁾ labeled with fluorescent dye JOE (Applied Biosystems), in a total volume of 50 μ l. Samples were denatured at 94°C for 4 min, followed by 27 cycles of 94°C for 1 min, 56°C for 1 min and 70°C for 1.5 min, with a final extension at 70°C for 5 min.

The hypervariable region 3' to the ApoB locus was amplified in essentially the same reaction buffer as D1S80 except that the relevant primers⁽⁸⁾ were labeled with the fluorescent dye FAM. Samples were amplified through 26 cycles at 94°C for 1 min, followed by 58°C for 6 min.

An ApoB allelic ladder was generated

by pooling a collection of amplified products that in combination represented approximately alternate bands throughout the allele size range. A 10^{-6} dilution of this cocktail was amplified as before, except primers labeled with the dye JOE were utilized.

Locus D17S5 was amplified in a reaction mix containing 1.5 mM each dNTP, 5 units of *Taq* polymerase, 6.7 mM $MgCl_2$, 6.7 mM Tris-HCl, 16.6 mM $(NH_4)_2SO_4$, 6.8 μM EDTA, 2 mM DTT, 10% DMSO, 5 μg BSA, and 0.1 μM each primer⁽⁶⁾ labeled with the dye TAMRA, in 50 μl total volume. Samples were denatured at 94°C for 5 min, followed by 27 cycles of 94°C for 0.5 min, 53°C for 0.5 min, and 65°C for 4 min, with a final extension at 65°C for 7 min. A minority of the alleles are greater than 1 kb in size, and thus products were verified by electrophoresing 20- μl aliquots for 2 hr at 80 V in a 1.0% SeaKem (FMC BioProducts) agarose gel and visualized by ethidium bromide staining/UV light.

Automated Product Analysis

Amplification products from different loci of the same individual were pooled and analyzed on an Applied Biosystems 362A as follows: 1 μl of each PCR product from ApoB, D17S5, and D1S80 was combined with 6 fmoles of internal lane standard comprising *AluI* digest fragments of pBR322 labeled with the dye ROX, then loaded in a 2% SeaPlaque (FMC BioProducts) agarose gel at a distance of 4 cm from the point of detection. Electrophoresis was for 5.5 hr at 100 V with 24 samples loaded per run; the same gel was reused up to five times. A single large D17S5 amplification product was sized by electrophoresing against a ROX-labeled *PstI* digest of bacteriophage λ as a standard, using a 1.8% SeaPlaque gel.

The fragment sizes were automatically determined by the software using the method of second-order regression to establish a curve of best fit generated from the internal standard in each lane.

Determination of Accuracy and Precision

The accuracy and precision of band size estimation by the 362A GeneScanner were determined for all three loci. These were calculated for all allele sizes using the following equations:

%Accuracy =

$$\left[1 - \frac{\sqrt{(\bar{x}/\text{Obs} - \text{Exp})^2}}{\text{Exp}} \right] \times 100$$

where \bar{x} Obs = average observed band size and Exp = theoretical allele size determined from previously published sequence data.^(7,8,17) This was calculated for each allele and an overall value was determined for each locus by calculating the weighted average.

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

Where SD = standard deviation of observed band sizes for a given allele and Exp = theoretical allele size.

RESULTS

An example of a gel picture generated by the coanalysis of the three loci in 24 unrelated individuals is shown in Figure 1. The shortest bands are at the bottom of the picture, with the internal size ladder shown in red and amplification products from loci D17S5, D1S80, and ApoB in yellow, green, and blue, respectively. An electrophoretogram generated from one of these samples is shown in Figure 2. The electrophoretogram depicts the fluorescence intensity of each band as it passes by the detector, with the shortest band at the left side of the figure. The calculated sizes of the VNTR bands are given in the table below the electrophoretogram (third column). The other bands in the electrophoretogram represent the size ladder.

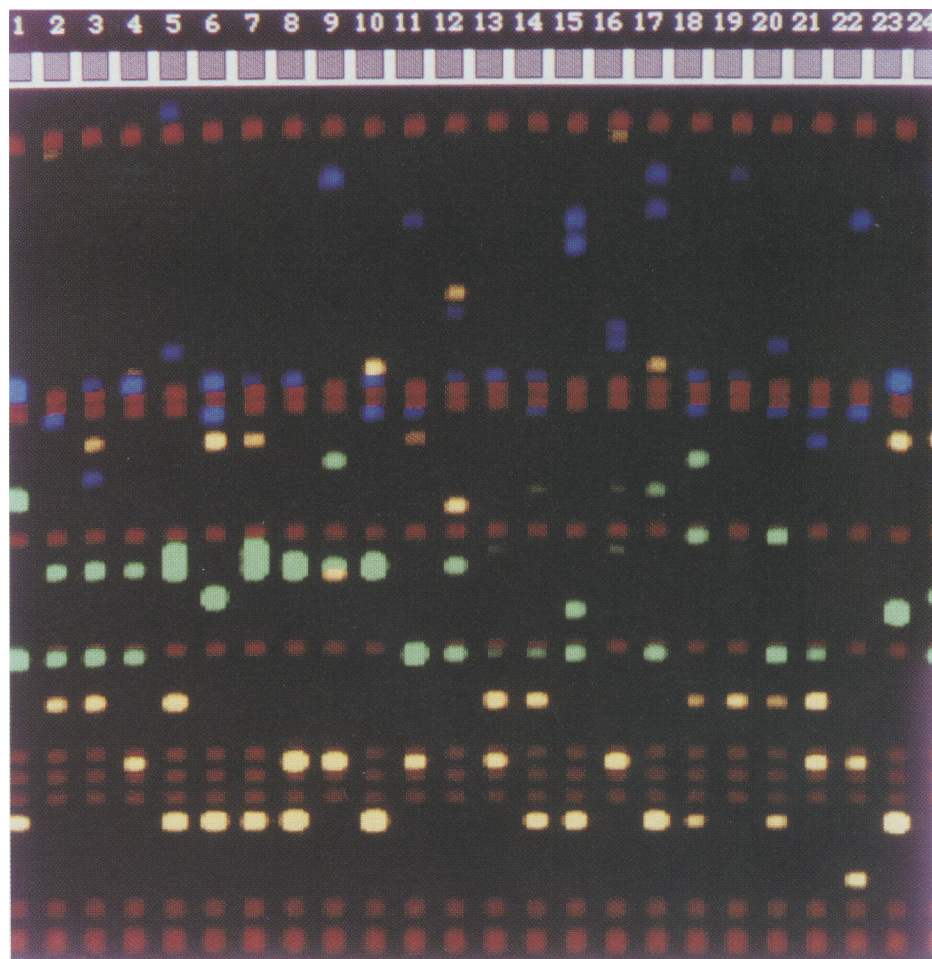


FIGURE 1 Genescanner gel file of electrophoresed PCR products from 24 unrelated individuals. D17S5, D1S80, and ApoB amplification products are shown as yellow, green, and blue bands, respectively, and the red bands are the internal standard of pBR322 restriction fragments. The smallest fragments are shown at the bottom of the picture.

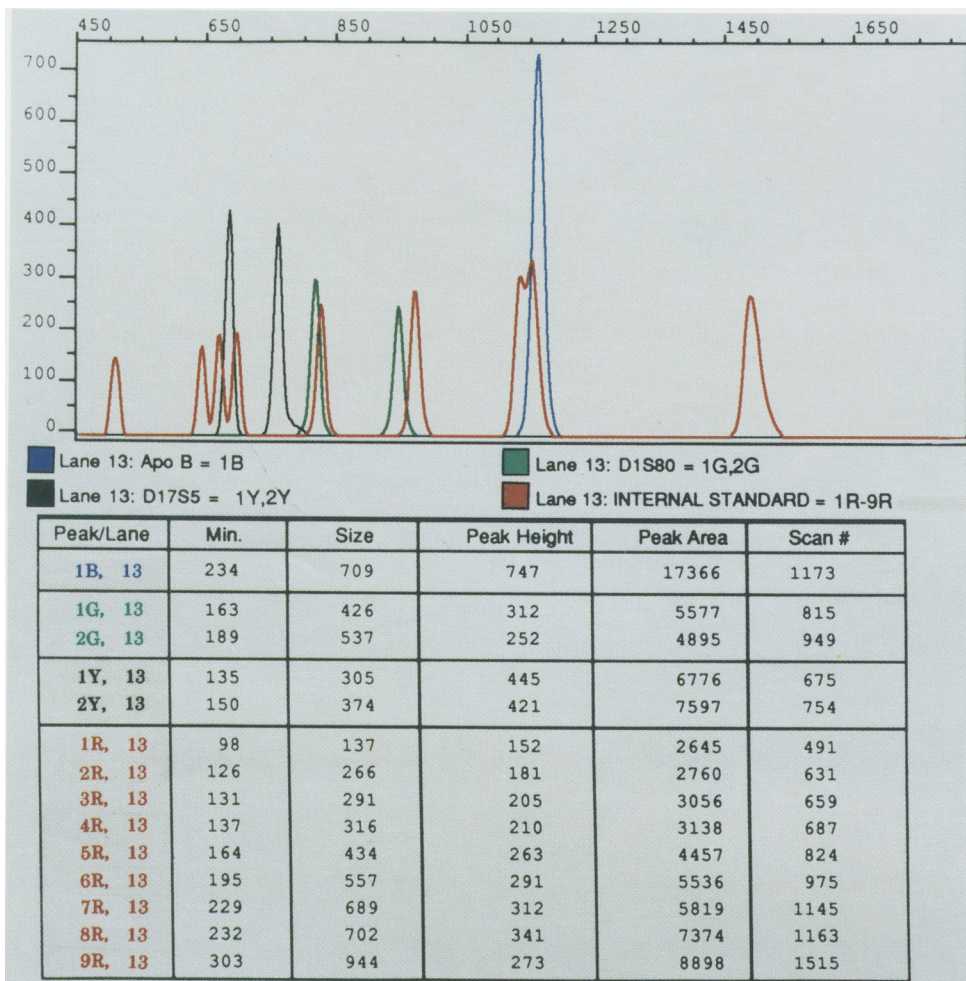


FIGURE 2 Electrophoretogram generated by the coanalysis of three VNTR loci. D17S5, D1S80, and ApoB amplification products are depicted as black, green, and blue peaks; the red peaks are the internal standard of pBR322 restriction fragments. Size in base pairs, plus peak height and area in arbitrary units, is determined for each fragment. Time (min) and scan number of band detection from the start of electrophoresis are also given.

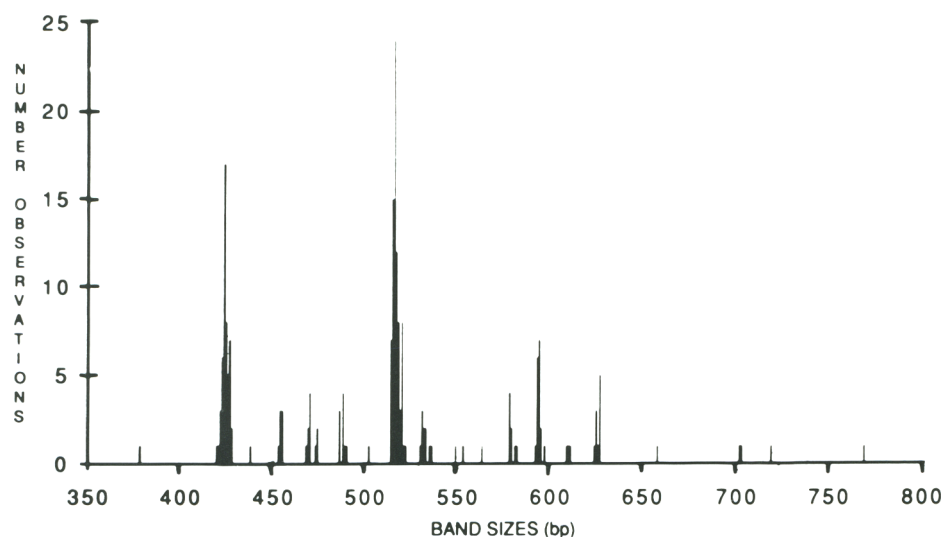


FIGURE 3 Frequency distribution of D1S80 allele sizes determined by automated fluorescence analysis of PCR products from a population of 107 unrelated British caucasians.

For locus D1S80, the fragment sizes fell in 19 distinct groups (maximum range 9 bp), defining the allele sizes, ~16 bp apart. Heterozygotes in which the two alleles differed by one repeat were readily typed, and analysis of reamplified samples gave identical allele assignments (data not shown). Frequency data was collected for 107 unrelated individuals (Fig. 3) and binned (i.e., arbitrarily grouped or pooled) to generate allele frequency data (see Fig. 6A). Because the precision of the data was very high (see below), the boundaries of the bins could be estimated realistically without resorting to a formal statistical analysis of the groups of similarly sized bands. For locus D17S5, 106 British caucasians were analyzed and alleles were detected in 15 discrete groups with a maximum size range of 6 bp for those samples falling within the molecular weight range of the *Alu*/pBR322 internal standard (i.e., up to 946 bp) (Fig. 4). Allele frequencies generated by binning each group are given in Figure 6B (below).

For ApoB, amplification products from 104 individuals were analyzed against both the pBR322/*Alu*I standard and the ApoB allelic ladder (Fig. 5A,B). Greater precision was obtained using the allelic ladder, and 15 discrete groups were discernible at ~30-bp spacing with a minority of bands falling between these groups. Allele frequencies were generated from this data (Fig. 6C). To investigate factors affecting the precision of ApoB band size estimation, we found that reamplifying ApoB loci did not alter the size-calling of the bands when the new products were electrophoresed in the same gel as aliquots of the original amplifications, regardless of the size standard used. Hence, the reamplification process was not causing detectable additions or deletions to the PCR products. Similarly, rerunning several aliquots of a single amplification product in the same gel gave identical calculated band sizes. However, aliquots of amplification products run on different 2% agarose gels were sized very similarly when compared with the ApoB allelic ladder but displayed significant variation when sized against the internal pBR322 digest (results not shown). From these experiments, we attribute the apparent difference in band sizes to anomalous migration rates for the ApoB repeats, which are highly rich in AT residues, relative to the migration rates

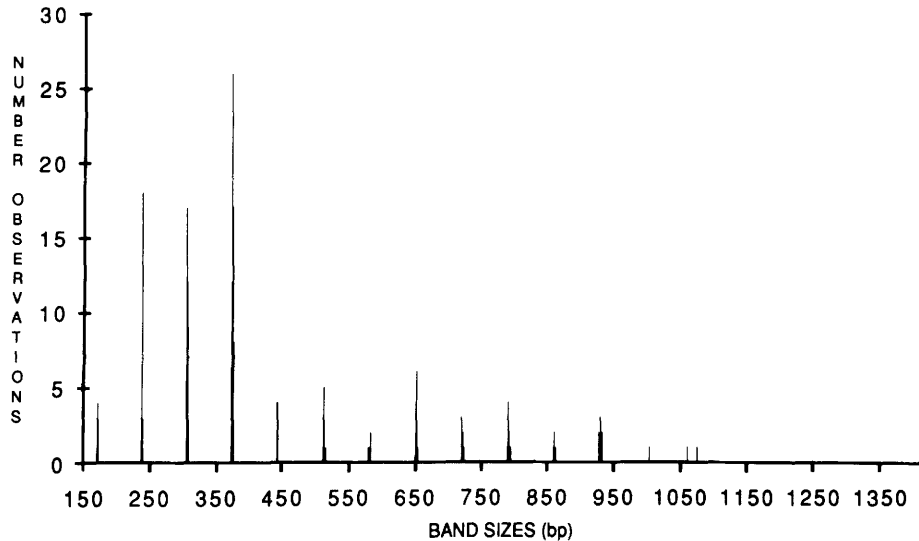


FIGURE 4 Frequency distribution of D17S5 allele sizes determined by automated fluorescence analysis of PCR products from a population of 106 unrelated British caucasians.

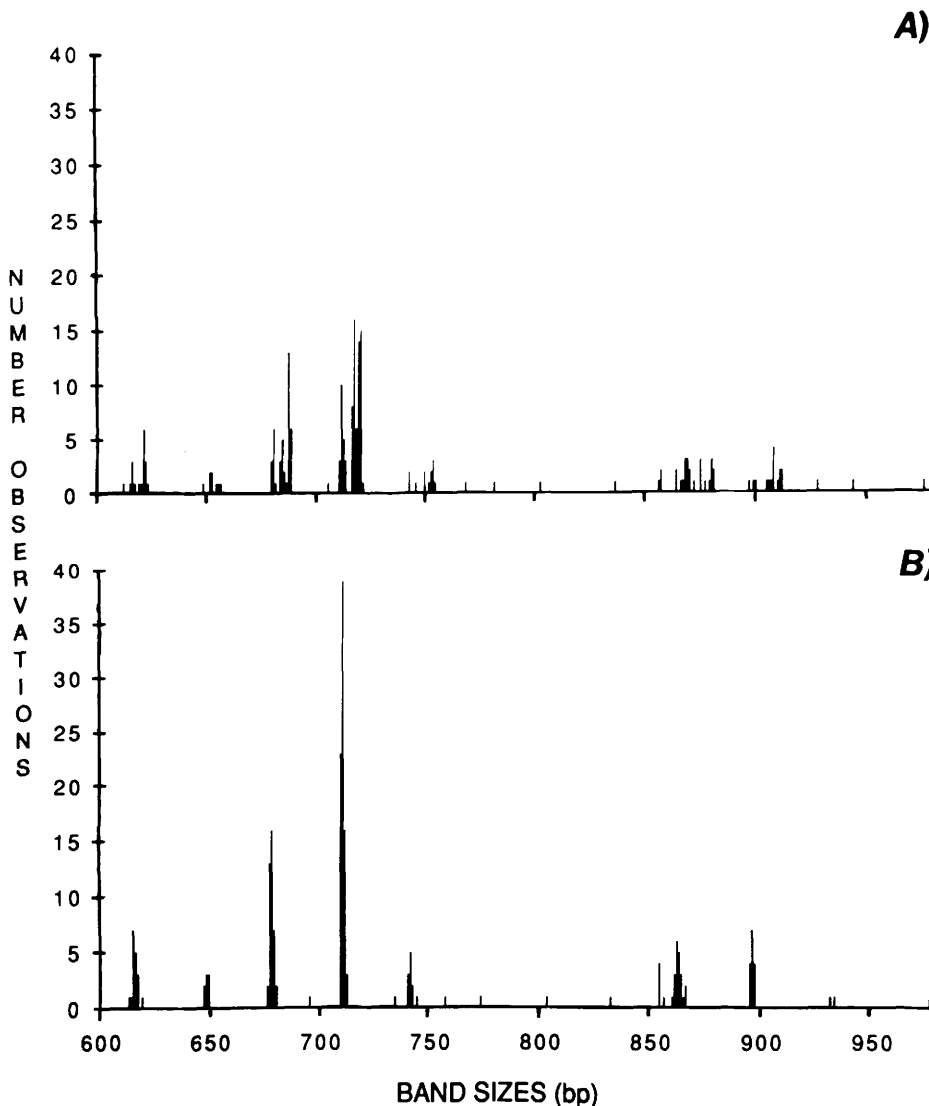


FIGURE 5 Comparison of ApoB allele sizes from 104 unrelated British caucasians determined using two different internal size standards: (A) pBR322/*AluI* standard and (B) ApoB allelic ladder.

of the fragments of the restriction digest of the pBR322, which are of random base composition.

The results for the combined discrimination power of the three loci are summarized in Table 1. A summary of the precision and accuracy of band size estimation is given in Table 2.

DISCUSSION

This study illustrates the feasibility of simultaneously characterizing the amplification products from three different VNTR loci in the same electrophoretic lane, using the dye-detection technology of the ABI362A GeneScanner. Because the overall length of the amplified alleles is small, even bands differing by a single repeat can be distinguished, and this enables the alleles to be defined according to the number of repeat units they contain. A major advantage to analyzing these loci is that different laboratories can directly compare their results regardless of the primers used or the technique employed to detect the amplification products. Their use is being extensively evaluated by the forensic laboratories of several European countries for whom the ability to compare results accurately is essential in the event of cross-border crime.

The electrophoretic run of 5.5 hr enabled all but one of the amplified alleles to be detected: Most D17S5 alleles comprise 1–14 copies of the 70-bp repeats, but rare large alleles do exist: A 30-repeat allele has been recorded by other workers in a survey of U.S. caucasians using SLP analysis.⁽¹⁸⁾ Thus, care should be taken in analyzing D17S5 by PCR to ensure that the larger alleles are amplified and detected whenever they are present.

We routinely include with each batch of D17S5 amplifications a positive DNA control comprising a 19-repeat allele as a quality check of the PCR reagents and conditions. We observed 88% heterozygosity at this locus within the British caucasian population by PCR/Genescanner analysis. This did not differ significantly from expected values assuming Hardy-Weinberg equilibrium and was similar to observed heterozygosity values for U.S. caucasians (86%) determined by RFLP analysis.⁽¹⁸⁾ Thus, the problem of dropout of high-molecular-weight alleles was not evident, because if significant numbers of alleles too large to be detected by amplification were

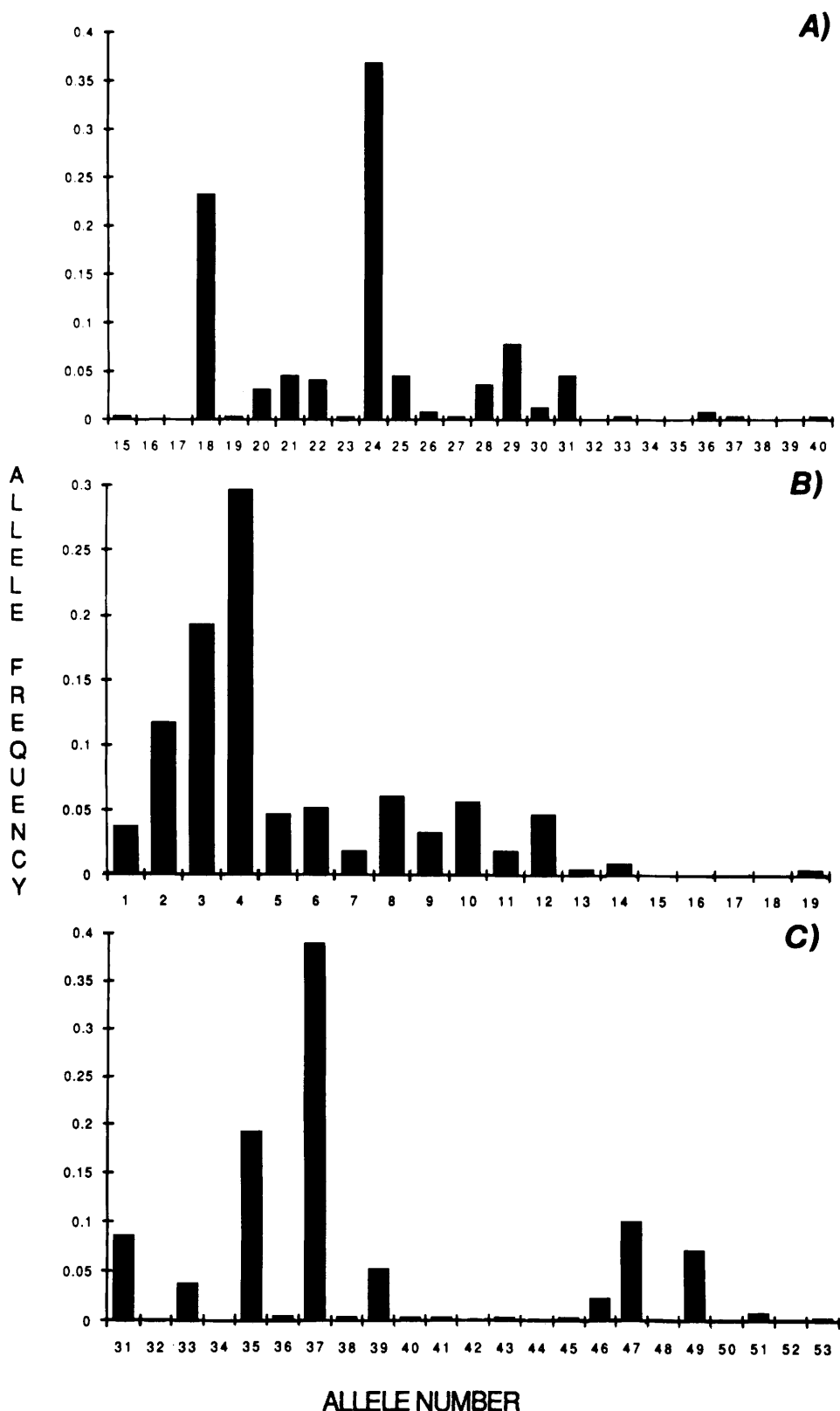


FIGURE 6 Allele frequency distributions for loci D1S80 (A), D17S5 (B), and ApoB (C) within the British caucasian population. These were generated by binning band size data from Figs. 2, 3, and 4B, respectively. Allele nomenclature is based on number of tandem repeats. Average sizes of commonest observed alleles were 517 bp (24 repeats) for D1S80, 374 bp (4 repeats) for D17S5, and 710 bp (37 repeats) for ApoB.

present, an excess of homozygotes would have been observed. However, it should be noted that these results were obtained with good-quality DNA; forensic samples can sometimes be partially degraded. Hence, it would be useful to include a DNA quality marker in the PCR reaction mix such as a coamplified invariant segment of human chromosomal DNA, which is bigger than the largest expected VNTR allele. Without such precautions, there is a risk of inadvertently scoring a partially degraded heterozygote as an apparent homozygote if the two alleles differ widely in size.

Under the specified amplification conditions, no spurious bands such as heteroduplexes or additional PCR products were generated at the three VNTR loci. Such artifacts can occur in VNTR amplification if the number of PCR cycles is excessive for a given initial concentration of target DNA template. Under such circumstances, another aliquot of the target DNA should be amplified using less template or fewer PCR cycles. For instance, with locus D1S80, amplification through 27 cycles gave clean results with 1–100 ng of template, but 23 cycles were optimal for 1 μ g initial target DNA (results not shown).

The size distribution of the ApoB alleles confirmed previous observations that these are comprised predominantly of 2×15 bp repeats.^(8,19) Anomalous migration rates of the ApoB alleles relative to an internal size marker could have been due to a DNA–gel interaction that is affected by slight differences in gel conformation, which in turn may be related to a DNA-bending type phenomenon to which adenine–thymine repetitive sequences are susceptible⁽²⁰⁾: The ApoB repeats comprise 97% A/T, 3% G/C,⁽⁸⁾ whereas pBR322 is 53% G/C. In contrast, D17S5 (64% G/C within the repeat) amplification products that were coelectrophoresed in the same gels did not display significant gel-dependent migration rates relative to the internal standard. However, scoring ApoB amplification products on agarose gels against an internal allelic ladder appeared to compensate for anomalies in migration rate by decreasing the average standard deviation in size calling to 0.95 bp (from 3.59 bp using pBR322/*Alu*I). This anomaly in migration rate is the subject of further investigation.

With loci D1S80 and D17S5, no bands

TABLE 1 Discrimination Power of Amplified VNTR Loci

Locus	Repeat unit length (bp)	Number of observed allele sizes	Probability of random match (pM)	Commonest genotype
D17S5	70	15	0.040	0.11
D1S80	16	19	0.065	0.17
ApoB	15	16	0.069	0.15
		Combined:	1.8×10^{-4}	2.8×10^{-3}

Probability of a random match (pM) was calculated from the equation $pM = \sum P_i^2$, where P_i is the genotype frequency.

TABLE 2 Accuracy and Precision of Band Size Estimation

Locus	Size standard	Percent accuracy	Percent precision (SD)
D17S5	pBR322/ <i>Alu</i> I	99.19	99.75 (1.02 bp)
D1S80	pBR322/ <i>Alu</i> I	97.80	99.67 (1.63 bp)
ApoB	pBR322/ <i>Alu</i> I	96.32	96.72 (3.59 bp)
ApoB	allelic ladder	97.29	99.85 (0.95 bp)

were recorded that fell midway between the respective allelic groups, indicating that adequate precision was afforded by the pBR322/*Alu*I size standard. However, if rare alleles exist that have fractions of repeat units or insertions/deletions of several base pairs within the amplified nonrepeat flanking regions, distinguishing these from the common bands may require the inclusion of appropriate allelic ladders to improve further the precision of band size estimation.

The combined probability of a random match at loci D1S80, D17S5, and ApoB is 1.8×10^{-4} for British caucasians and the frequency for the combined most common genotype is 1 in 300 (Table 1). The test compares favorably with the Amplitype HLA DQ alpha Forensic DNA Amplification and Typing Kit (Cetus Corporation, Emeryville, CA), currently the most extensively reported forensic PCR test, which in caucasians has a probability of a random match of 0.07 plus commonest genotype frequency of ~ 1 in 8.⁽²¹⁻²³⁾ Thus, the test has applicability in all fields of human identity testing such as forensic analysis and paternity testing. Multiplex amplification of the loci would further increase the automation and efficiency of the combined VNTR analysis test. In our hands, ApoB and D1S80 coamplify under the conditions stated earlier for D1S80, but these are not compatible with D17S5 amplification. Multiplex amplification of small

VNTRs will become easier as more loci that have similar optimal amplification conditions become available. Unfortunately, only four distinguishable fluorescent tags are currently available, but this limitation can be overcome if the same dye tag is used on loci with mutually exclusive allele size ranges such as in the multiplex analysis of microsatellite loci, which typically display restricted size spans.

The results show that the accuracy and precision of the fluorescence measurements are sufficient to enable VNTRs to be correctly and reproducibly characterized by the automated technique.

REFERENCES

- Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeats (markers) for human gene mapping. *Science* **235**: 1616-1622.
- Gill, P., A.J. Jeffreys, and D.J. Werrett. 1985. Forensic application of DNA fingerprints. *Nature* **318**: 577-579.
- Jeffreys, A.J., V. Wilson, R. Neumann, and J. Keyte. 1988. Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16**: 10953-10971.
- Wong, Z., V. Wilson, A.J. Jeffreys, and S.L.

- Thien. 1986. Cloning a selected fragment from a human DNA fingerprint: Isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* **14**: 4605-4616.
- Wong, Z., V. Wilson, I. Patel, S. Povey, and A.J. Jeffreys. 1987. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**: 269-288.
- Horn, G.T., B. Richards, and K.W. Klinger. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucleic Acids Res.* **17**: 2140.
- Kasai, K., Y. Nakamura, and R. White. 1990. Amplification of a variable number tandem repeat (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *J. Forensic Sci.* **5**: 1196-1200.
- Boerwinkle, E., W. Xiong, E. Fourest, and L. Chan. 1989. Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: Application to the apolipoprotein B 3' hypervariable region. *Proc. Natl. Acad. Sci.* **86**: 212-216.
- Allen, R.C., G. Graves, and B. Budowle. 1989. Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. *BioTechniques* **7**: 736-744.
- Carrano, A.V., J. Lamerdin, L.K. Ashworth, B. Watkins, E. Brankscomb, T. Slezak, M. Raff, P. de Jong, D. Keith, L. McBride, S. Meister, and M. Kronick. 1989. A high-resolution, fluorescence based, semi-automated method for DNA fingerprinting. *Genomics* **4**: 129-136.
- Chehab, F.F. and Y.W. Kan. 1989. Detection of specific DNA sequences by fluorescence amplification: A color complementation assay. *Proc. Natl. Acad. Sci.* **86**: 9178-9182.
- Mayrand, P.E., J. Robertson, J. Ziegler, L.B. Hoff, L.J. McBride, J.S. Chamberlain, and M.N. Kronick. 1991. Automated genetic analysis. *Ann. Biol. Clin.* **49**: 224-230.
- Chehab, F.F. and Y.W. Kan. 1990. Detection of sickle cell anaemia mutation by colour DNA amplification. *Lancet* **335**: 15-17.
- Landgraf, A., B. Reckmann, and A. Pingoud. 1991. Quantitative analysis of polymerase chain reaction (PCR) products using primers labeled with biotin and a fluorescence dye. *Anal. Biochem.* **193**: 231-235.
- Robertson, J.M., J. Ziegler, M. Kronick, D. Madden, and B. Budowle. 1991. Genetic typing using automated electrophoresis and fluorescence detection. In *DNA fingerprinting: Approaches and applications* (ed. T. Burke, A. Dolf, A.J. Jeffreys, and R. Wolff), pp. 391-398. Birkhaeuser Verlag, Basel.
- Gill, P., C.P. Kimpton, and K.M. Sullivan. 1992. A rapid method for identifying

- fixed specimens by DNA profiling. *Electrophoresis* **13**: 173–175.
17. Wolff, R.K., Y. Nakamura, and R. White. 1988. Molecular characterization of a spontaneously generated new allele at a VNTR locus: No exchange of flanking DNA sequence. *Genomics* **3**: 347–351.
 18. Odelberg, S.J., R. Plaetke, J.R. Eldridge, L. Ballard, P. O'Connell, Y. Nakamura, J.M. Leppert, and R. White. 1989. Characterization of eight VNTR loci by agarose gel electrophoresis. *Genomics* **5**: 915–924.
 19. Ludwig, E.H., W. Friedl, and J. McCarthy. 1989. High-resolution analysis of a hypervariable region in the human apolipoprotein B gene. *Am. J. Hum. Genet.* **45**: 458–464.
 20. Hagerman, P.J. 1990. Sequence directed curvature of DNA. *Annu. Rev. Biochem.* **59**: 755–781.
 21. Helmuth, R., N. Fildes, E. Blake, M.C. Luce, J. Chimera, R. Madej, C. Gorodezky, M. Stoneking, N. Schmill, W. Klitz, R. Higuchi, and H.A. Erlich. 1990. HLA-DQ α allele and genotype frequencies in various human populations, determined using enzymatic amplification and oligonucleotide probes. *Am. J. Hum. Genet.* **47**: 515–523.
 22. Comey, C.T.R. and B. Budowle. 1991. Validation studies on the analysis of the HLA DQ alpha locus using the polymerase chain reaction. *Forensic Sci.* **36**: 1633–1648.
 23. Sullivan, K.M., P. Gill, D. Lingard, and J.E. Lygo. 1992. Characterization of HLA DQ alpha for forensic purposes. Allele and genotype frequencies in British Caucasian, Afro-Caribbean and Asian populations. *Int. J. Leg. Med.* (in press).

Received December 30, 1991; accepted in revised form May 21, 1992.