



Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples.

P Pacek, A Sajantila and A C Syvänen

Genome Res. 1993 2: 313-317

Access the most recent version at doi:[10.1101/gr.2.4.313](https://doi.org/10.1101/gr.2.4.313)

References This article cites 15 articles, 3 of which can be accessed free at:
<http://genome.cshlp.org/content/2/4/313.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

Determination of Allele Frequencies at Loci with Length Polymorphism by Quantitative Analysis of DNA Amplified from Pooled Samples

Päivi Pacek, Antti Sajantila, and Ann-Christine Syvänen

Department of Human Molecular Genetics, National Public Health Institute, SF-00300 Helsinki, Finland

We present a new method that allows rapid determination of allele frequencies at loci exhibiting length polymorphism. In this method a fluorescence-labeled PCR primer is used to amplify the polymorphic region from pooled DNA samples originating from a large number of individuals. The fluorescent PCR products are separated by gel electrophoresis on an automatic DNA sequencer and the relative amount of the PCR products are determined. The distribution of the PCR products obtained from the alleles present in the pooled samples directly corresponds to the allele frequency in the population in question. The allele frequencies at a short tandem repeat locus in the von Willebrand factor gene and at the D1S80 locus were determined in the Finnish population. We found that the allele frequencies determined by quantitative analysis of PCR products from pooled DNA samples and by analyzing individual samples were in good agreement.

Analysis of polymorphic dinucleotide repeat,⁽¹⁾ short tandem repeat (STR)⁽²⁾ or variable number of tandem repeat (VNTR) regions⁽³⁾ of the human genome have proven to be highly useful in forensic identification and in genetic linkage studies of human diseases. The polymorphic alleles can be conveniently identified with the aid of the PCR⁽⁴⁾ using primers flanking the repeated core unit. The amplified alleles are detected most commonly after gel electrophoretic size separation by using autoradiography, staining with ethidium bromide,^(5,6) or silver nitrate.⁽⁷⁾ A fluorescent label can also be introduced into the amplification product by mediation of the PCR primer, which allows the alleles to be detected during the electrophoresis with a fluorescence detector.⁽⁸⁻¹⁰⁾

The statistical interpretation of the results of forensic and linkage analyses requires information on the population frequency of the specific alleles at each locus analyzed. To determine accurately the allele frequencies at loci revealing multiple alleles, a large number of samples in the population of interest have to be analyzed.

Here, we introduce a new general approach that allows rapid determination of the allele frequencies at loci exhibiting length polymorphism. The method is based on amplification of the polymorphic region from a pooled sample containing DNA from a large number of individuals using a fluorescence-labeled PCR primer. The PCR products synthesized from each allele are separated and quantified using an automatic fluorescence sequencing instrument. The distri-

bution among the PCR products obtained from the alleles in the DNA pool corresponds well to the frequencies of the alleles in the population sample.

In the present study we applied this method to determine the Finnish allele frequencies at a STR locus in the von Willebrand factor gene (VWA) on chromosome 12⁽¹¹⁾ and at the D1S80 (MCT118) VNTR locus on chromosome 1.⁽⁵⁾ At the VWA locus the polymorphism is caused by a tetranucleotide repeat unit. At the D1S80 locus, where the polymorphism is the result of a 16-bp tandem repeat unit, a distribution of 15 alleles has been reported previously in the Finnish population.⁽¹²⁾

MATERIALS AND METHODS

DNA Samples

Batches of leukocytes, which had been used previously in the production of interferon- α ⁽¹³⁾ were obtained from the Finnish Red Cross Blood Transfusion Service. Each batch originated from 400 ml of blood from a known number (between 90 and 200) of donors living in the capital region of Finland. DNA was purified from an aliquot of each leukocyte batch according to a standard method.⁽¹⁴⁾ The concentration of the DNA was determined by measuring the absorbance at 260 nm. The DNA isolated from the leukocytes was combined into larger pools, taking into account the number of individuals in each batch, to yield four pools containing an equal amount of DNA from 1350, 690, 1100, and 180 individuals, respectively. DNA was also purified from individual blood

samples,⁽¹⁴⁾ which had been sent for disputed paternity analysis to the National Public Health Institute, Helsinki, Finland.

Primers

The primers for amplification of the VWA locus were those described by Kimpton et al.,⁽¹⁵⁾ and they had the following sequences: 5'-CCC TAG TGG ATG ATA AGA ATA ATC (upstream primer; nucleotide 1640–1663 of intron 40) and 5'-GGA CAG ATG ATA AAT ACA TAG GAT GGA TGG (downstream primer; nucleotide 1793–1764 of intron 40). The size of the PCR product obtained with these primers is 150–180 bp. The D1S80 primers were those described by Kasai et al.⁽⁵⁾ Their sequences were 5'-GAA ACT GGC CTC CAA ACA CTG CCC GCC G (upstream primer) and 5'-GTC TTG TTG GAG ATG CAG GTG CCC CTT GC (downstream primer). The size of the PCR product obtained with these primers is 430–750 bp.

The primers were synthesized on an Applied Biosystems 392 DNA synthesizer. The 5' ends of the upstream VWA and the downstream D1S80 primer were fluorescence-labeled during the synthesis using the FluorePrime reagent (Pharmacia LKB Biotechnology AB). The primers were used without purification.

PCR

One hundred (30,000 molecules), 10 (3,000 molecules), or 1 ng (300 molecules) of DNA was amplified using 20 nmoles of the four dNTPs, 100 pmoles of each primer, and 2.5 units of *Taq* DNA polymerase (Promega Biotech) in 100 μ l of 50 mM Tris-HCl (pH 8.8), 15 mM $(\text{NH}_4)_2\text{SO}_4$, 1.5 mM MgCl_2 , 0.1% Triton X-100, and 0.01% gelatin. The PCR was initiated with a "hot start" by first heating the samples for 5 min at 95°C, followed by the addition of the *Taq* DNA polymerase at 80°C. Thirty PCR cycles of 1 min at 95°C, 1 min at 58°C, and 1 min at 72°C were carried out with the VWA primers, and 30 cycles of 1 min at 95°C, 1 min at 65°C, and 5 min at 72°C were carried out with the D1S80 primers.

Quantitative Analysis of the PCR Products

The fluorescent PCR products were separated on 6% denaturing polyacrylamide

gels using an automatic DNA Sequencer (A.L.F., Pharmacia LKB Biotechnology AB). Between 0.25 and 5 μ l of the PCR products were analyzed after denaturation at 95°C for 3 min in 50% formamide containing blue dextran. The samples were run at 1500 V for 3 hr (VWA) or 8 hr (D1S80) at 40°C. The relative amount of PCR product obtained in each sample was determined with the aid of the A.L.F. DNA Fragment Manager V1.0 program, kindly provided to us by Per Johan Ulfendahl and Margareta Degerman (Pharmacia LKB, Molecular Biology Systems Division, Uppsala, Sweden).

Amp-FLP Analysis

The individual DNA samples were typed by the amplified fragment length polymorphism (Amp-FLP) technique⁽⁷⁾ using vertical polyacrylamide gels and staining with silver nitrate as described previously.⁽¹²⁾

RESULTS

The method described here is based on quantitative analysis of PCR products obtained from alleles varying in length. Because the efficiency of the PCR can be expected to depend on the size of the amplified allele, we first determined the relative amount of the two PCR products obtained from the DNA of six heterozygous individuals with different combinations of the VWA alleles. In these samples the two alleles differed in size by from 1 to 6 tetranucleotide repeat units, and the size of the PCR products varied between 150 and 180 bp. Comparison of the amount of PCR product from both alleles in each sample showed that the efficiency of the amplification of the alleles is independent of their size difference (Table 1). The amount of DNA (100, 10, or 1 ng) subjected to the PCR did not affect the allele distribution (data not shown).

The alleles at the VWA locus were then amplified from three pooled samples containing an equal amount of DNA from 1,350, 690 and 1,100 Finnish individuals, respectively, and the amount of PCR product obtained from each allele in the pooled sample was determined. Figure 1 shows the result of the electrophoretic separation of the alleles amplified from the sample containing DNA from 1100 individuals as registered by the fluorescence detector of the sequenc-

TABLE 1 Relative Amounts of PCR Products Obtained from the VWA Alleles of Individuals with Different Genotypes

| Sample | Allele combination ^a | Relative amount of PCR product ^b |
|--------|---------------------------------|---|
| W1 | 1–7 | 0.50–0.50 |
| W2 | 1–6 | 0.47–0.53 |
| W3 | 1–5 | 0.55–0.45 |
| W4 | 2–5 | 0.50–0.50 |
| W5 | 3–5 | 0.47–0.53 |
| W6 | 3–4 | 0.50–0.50 |

^aThe alleles are designated according to their size, with the smallest allele being number 1.

^bMean values of three determinations.

ing instrument. The area of the peaks in the curve reflects the relative amount of PCR product from each allele. Figure 2 shows the allele frequencies obtained by determining the area of the peaks in each curve. Similar allele distributions were obtained from the three samples. The same result was obtained when 100, 10, or 1 ng of the pooled DNA was amplified (data not shown).

We also determined the allele frequencies at the VWA locus from 100 individual samples using the Amp-FLP technique. The allele frequencies determined from individual and from pooled leukocytes were in good agreement (Table 2).

Table 3 shows the distribution of the PCR products obtained by amplification of the alleles at the D1S80 locus from DNA of eight heterozygous individuals with different allele combinations. Contrary to the result at the VWA locus, at the D1S80 locus we found that the smaller allele was amplified more efficiently than the large one in all samples, but the difference in efficiency was small for alleles that differed from each other by only one 16-bp repeat unit. The amount of DNA (100 or 10 ng) subjected to the PCR did not affect the result (data not shown).

Because the amplification efficiency did not vary significantly between adjacent alleles, we proceeded to determine the allele frequencies at the D1S80 locus from two pooled DNA samples representing 690 and 180 individuals, respectively. Figure 3 shows the electrophoretic separation of the alleles amplified from the samples containing DNA from 690 individuals. Table 4 com-

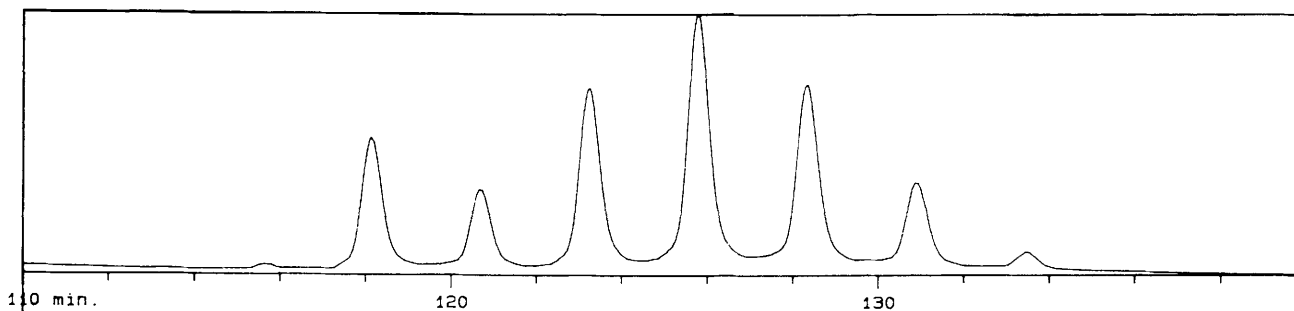


FIGURE 1 Separation of the VWA alleles amplified from 10 ng of a pooled sample containing DNA from 1100 individuals on the automatic DNA sequencer.

compares the allele frequencies determined from the pooled samples to those determined previously in the Finnish population from 140 individual samples by the Amp-FLP technique.⁽¹²⁾ The correlation between the two methods is good, even though a tendency toward higher frequencies of the smaller alleles and lower frequencies of the larger alleles in the pooled samples can be noticed. Two rare alleles observed in the individual samples remained below the detection limit in the pooled samples.

DISCUSSION

We describe a novel strategy to determine the allele frequencies in the population at loci exhibiting length polymorphism. We used quantitative analysis of

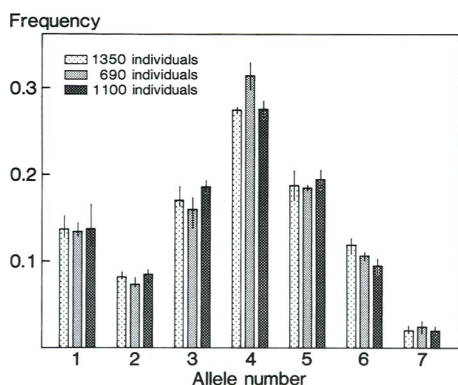


FIGURE 2 Allele frequencies at the VWA locus determined by amplification of the DNA from three pooled samples (1350, 690, and 1100 individuals, respectively). The PCR products were separated on the DNA sequencer, and the relative amount of PCR products was determined using the A.L.F. DNA Fragment Manager Program V1.0. The columns show the mean values obtained from three separate PCRs from each pool. The variation range between the determinations are indicated by vertical lines.

PCR products obtained from pooled DNA samples representing a large number of individuals to determine the allele frequencies at two loci. This approach greatly simplifies the determination of allele frequencies because the same information that is obtained by analyzing hundreds of individual samples is obtained in one or a few reactions. One disadvantage of our method, compared with analysis of individual samples, is that the distribution of genotypes in the population can only be deduced by assuming that the marker conforms to Hardy-Weinberg expectations.

In this study the relative amounts of the amplified alleles were determined by fluorescence detection using an automatic DNA sequencer. The PCR products are labeled by mediation of a fluorescent primer, and consequently, each molecule of PCR product carries one molecule of label. Contrary to the detection of PCR products by staining gels with ethidium bromide or silver nitrate, where the signals depend on the size of the ampli-

fied fragment, the result of the fluorescent quantitative analysis is obtained as the molar distribution between the alleles.

At the VWA STR locus, where the alleles differ in size by only a repetitive tetranucleotide unit, and the size of the PCR products range from 150 bp to 180 bp, the method gave an accurate estimate of the allele frequencies as compared with the allelic frequencies determined from 100 individual samples. At the D1S80 locus, where the alleles differ by a 16-bp repeat unit and the size of the PCR products vary from 430 to 750 bp, we found that the smaller alleles were amplified slightly more efficiently, and, as could be expected the difference in PCR efficiency depended on the difference in size between the alleles. Despite this, the correlation between the allele frequencies determined from the pooled samples and those determined by analyzing individual samples was acceptable.

More efficient ("preferential") amplification of smaller VNTR alleles has been

TABLE 2 Comparison of Allele Frequencies at the VWA Locus Determined from Pooled and Individual Samples

| Allele | Allele frequency (samples) | |
|--------|----------------------------|-------------------------|
| | pooled ^a | individual ^b |
| 1 | 0.136 | 0.125 |
| 2 | 0.081 | 0.040 |
| 3 | 0.172 | 0.195 |
| 4 | 0.289 | 0.310 |
| 5 | 0.190 | 0.175 |
| 6 | 0.106 | 0.130 |
| 7 | 0.022 | 0.025 |

^aMean value of the frequencies given in Fig. 2, 3140 individuals are represented in the pools.

^bResults from 100 individual samples.

TABLE 3 Relative Amounts of PCR Products Obtained from the D1S80 Alleles of Individuals with Different Genotypes

| Sample | Allele combination ^a | Relative amount of PCR product ^b |
|--------|---------------------------------|---|
| M1 | 18–31 | 0.63–0.37 |
| M2 | 18–24 | 0.52–0.48 |
| M3 | 24–28 | 0.65–0.35 |
| M4 | 18–19 | 0.53–0.47 |
| M5 | 24–25 | 0.51–0.49 |
| M6 | 28–29 | 0.54–0.46 |
| M7 | 29–30 | 0.51–0.49 |
| M8 | 31–32 | 0.55–0.45 |

^aThe alleles are designated according to the number of repeat units.⁽¹²⁾

^bMean values of three determinations.

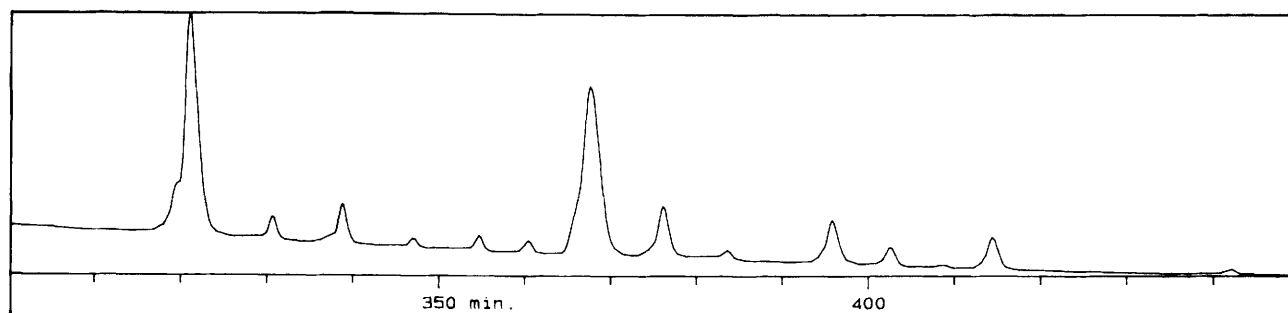


FIGURE 3 Separation of the D1S80 alleles amplified from 10 ng of a pooled sample containing DNA from 690 individuals on the DNA sequencer.

encountered particularly at the D17S30 (YNZ22) locus,^(6,16) where the size difference between the alleles is large (70 bp). It has been suggested that this problem can be avoided by raising the molar ratio of the *Taq* DNA polymerase to the template DNA during the PCR.⁽¹⁶⁾ We have also found that loss of the larger YNZ22 allele can be avoided by decreasing the amount of DNA template subjected to the PCR (A. Sajantila, M. Lukka, and A. C. Syvänen, unpubl.). In this study we found that the amount of DNA template subjected to PCR did not affect the relative distribution between the amplified alleles in the range tested at the VWA or D1S80 locus. The sensitive quantitative determination of PCR products using a DNA sequencer is a useful tool for further studies of the factors that affect the

TABLE 4 Comparison of the Allele Frequencies at the D1S80 Locus Determined from Two Pooled Samples and from Individual Samples

| Allele | Allele frequency | | |
|--------|-------------------|-------------------|---|
| | pool (n = 690) | pool (n = 180) | individual samples (n = 140) ^a |
| 18 | 0.343 | 0.331 | 0.307 |
| 19 | 0.024 | 0.014 | 0.011 |
| 20 | 0.048 | 0.034 | 0.032 |
| 21 | 0.010 | 0.008 | 0.018 |
| 22 | 0.017 | 0.020 | 0.014 |
| 23 | 0.012 | 0.008 | 0.014 |
| 24 | 0.330 | 0.343 | 0.311 |
| 25 | 0.073 | 0.071 | 0.075 |
| 26 | 0.010 | 0.004 | 0.011 |
| 27 | 0 | 0 | 0.007 |
| 28 | 0.059 | 0.047 | 0.068 |
| 29 | 0.022 | 0.039 | 0.032 |
| 30 | 0.006 | 0.009 | 0.043 |
| 31 | 0.042 | 0.067 | 0.079 |
| 37 | 0 | 0 | 0.007 |

^aFrom Sajantilla et al.⁽¹²⁾

efficiency of the PCR at loci exhibiting length polymorphisms.

From a practical point of view, an advantage of detecting PCR products by silver staining is that a lower efficiency of the amplification of larger alleles is compensated by more efficient staining of them. If scanning of silver-stained gels would be applied to determine allele frequencies from pooled samples, an overestimate of the frequency of larger alleles could be expected.

Our results showed that STR alleles differing in size by a tetranucleotide unit are amplified with equal efficiency, which indicates that STR markers with short alleles may be the markers of choice for PCR-based DNA typing. On the other hand, when using markers having a repeated core unit of only two nucleotides, a frequently encountered problem in PCR-based typing is artificial amplification products one or two repeat units smaller in size than the actual allele. These artifacts may account for a substantial proportion of the PCR products and, thus, hamper the DNA typing. In such cases, our method of determining allele frequencies from pooled samples would be difficult to apply. When analyzing individual samples at the VWA locus, we observed a small amount of PCR product one repeat unit smaller than the real alleles, but the amount of this product never exceeded 5%. When amplifying pooled samples we repeatedly obtained a small amount (<0.8%) of product one repeat unit smaller than allele 1 (Fig. 1). This was interpreted as being a PCR artifact but may also represent a rare allele.

In conclusion, we have devised a rapid procedure for determination of the population frequencies of alleles at polymorphic STR or VNTR loci. The number of reported polymorphisms in the human genome is increasing rapidly

at present. Therefore, this convenient method to evaluate the informativity of new markers to be applied to forensic identification or genetic linkage studies in a particular population should be highly useful.

ACKNOWLEDGMENTS

We are grateful to Hanna-Leena Kauppinen (The Finnish Red Cross Blood Transfusion Service) for providing the leukocyte pools, to Per Johan Ulfendahl and Margareta Degerman for giving us access to the A.L.F. DNA Fragment Manager V1.0 program, and to Bruce Budowle for the VWA primer sequences and for his professional comments on the manuscript. We thank Leena Karttunen, Tuula Manninen, Leena Peltonen, Sari Röksä, and Miikka Vikkula for their help during this study.

REFERENCES

1. Weber, J.L. and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388–396.
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. 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 repeat (VNTR) markers for human gene mapping. *Science* **235**: 1616–1622.
4. Mullis, K.B. and F.A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**: 335–350.
5. Kasai, K., Y. Nakamura, and R. White. 1990. Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reac-

- tion (PCR) and its application to forensic science. *J. Forensic Sci.* **35**: 1196–1200.
6. Horn, G.T., B. Richards, and K.W. Klinger. 1989. Amplification of a highly polymorphic VNTR segment by polymerase chain reaction. *Nucleic Acids Res.* **17**: 2140.
 7. Budowle, B., R. Chakraborty, A.M. Giusti, A.J. Eisenberg, and R. Allen. 1991. Analysis of the VNTR locus D1S80 by the PCR followed by high resolution PAGE. *Am. J. Hum. Genet.* **48**: 137–144.
 8. Robertson, J.M., J. Ziegle, 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.
 9. 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.
 10. Clemens, P.R., R.G. Fenwick, J.S. Chamberlain, R.A. Gibbs, M. de Andrade, R. Chakraborty, and C.T. Caskey. 1991. Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. *Am. J. Hum. Gen.* **49**: 951–960.
 11. Manusco, D.J., E.A. Tuley, L.A. Westfield, N.K. Worrall, B.B. Shelton-Inloes, J.M. Sorace, Y.G. Alevy, and J.E. Sadler. 1989. Structure for the gene for human von Willebrand factor. *J. Biol. Chem.* **264**: 19514–19527.
 12. Sajantila, A., B. Budowle, M. Ström, V. Johnsson, M. Lukka, L. Peltonen, and C. Ehnholm. 1992. PCR amplification of alleles at the D1S80 locus: Comparison of a Finnish and a North American Caucasian population sample, and forensic case-work evaluation. *Am. J. Hum. Genet.* **50**: 816–825.
 13. Cantell, K., S. Hirvonen, H.L. Kauppinen, and G. Myllylä. 1981. Production of interferon in human leukocytes from normal donors with the use of Sendai virus. *Methods Enzymol.* **78**: 29–38.
 14. Bell, G.I., J.H. Karam, and W.J. Rutter. 1981. Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Proc. Natl. Acad. Sci.* **78**: 5759–5763.
 15. Kimpton, C., A. Walton, and P. Gill. 1992. A further tetranucleotide repeat polymorphism in the vWF gene. *Hum. Mol. Genet.* **1**: 287.
 16. Walsh, P.S., H.A. Erlich, and R. Higuchi. 1992. Preferential PCR amplification of alleles: Mechanisms and solutions. *PCR Methods Applic.* **1**: 241–250.

Received October 30, 1992; accepted in revised form December 17, 1992.