Eur. J. Clin. Chem. Clin. Biochem.
Vol. 29, 1991, pp. 321–325

Individual-Specific DNA Fingerprinting in Man Using the Oligonucleotide Probe $(GTG)_5/(CAC)_5$

By *C. Peters*[1,3], *V. Schneider*[1], *J. T. Epplen*[2] and *H. Pöche*[1]

[1] *Institut für Rechtsmedizin, Freie Universität Berlin, Berlin (Dahlem), Germany*
[2] *Max-Planck-Institut für Psychiatrie, Martinsried, Germany*

(Received September 8, 1990/January 11, 1991)

Summary: Restriction fragment length polymorphisms (RFLPs) associated with interspersed simple repetitive DNA arise from DNA fragment lengths that contain variable numbers of the repeated motifs. Using restriction enzymes with different 4 base pair recognition sites and the simple triplet repeat hybridization probe, $(GTG)_5/(CAC)_5$, DNA multilocus fingerprints can be obtained in man. Only the DNAs of monozygous twins show indistinguishable banding patterns. Since the bands are inherited according to *Mendel*ian laws, DNA fingerprints can be used for identification of individuals and paternity analysis. The discriminatory power in the DNA fingerprinting technique in forensic science is demonstrated and examples of paternity testing are given.

Introduction

The human genome contains approximately 7×10^9 nucleotides per diploid cell, these nucleotides being distributed unequally between the 23 chromosome pairs. Some 20% to 30% of the human genome is comprised of repetitive sequences. The function of most of them is still not known. Many repeats are not transcribed significantly and the few that are, are not translated (1). Several repetitive sequences are known to be rich sources of inter-individual genetic variation (2), showing more diversity than the unique DNA sequence regions (3). Recently, studies on genetic variation in man, based on the analysis of so-called hypervariable loci, have attained great importance in medical science. Several hypervariable loci are composed of tandemly organized repetitive DNA sequences which tend to be hypervariable in copy number (4, 5). Alterations in the number of tandemly repeated DNA units or changes of their flanking restriction enzyme recognition sites produce DNA fragment length variations (Restriction Fragment Length Polymorphisms – RFLPs). The latter can be detected after digestion of a DNA sample with any frequently cutting restriction endonuclease that does not cut within the repeat unit. Subsequent hybridization with certain DNA probes detects many hypervariable loci simultaneously (6–10), and produces DNA banding patterns which show somatic stability. Two main categories of repetitive DNA have been used in DNA fingerprinting:

i) so called "minisatellites" which are tandemly repeated DNA sequences sharing a common "core" sequence (11, 12) of 16–33 base pairs; and

ii) "simple repetitive DNA sequences", i. e. short, tandemly repeated units of 2 to 10 pase pair long motifs (13).

A panel of simple repeat motifs has been examined in more than 120 animal and plant species (*Epplen*, unpublished results). Some of a large variety of clinical DNA fingerprinting applications are e. g. chimaera analysis after bone marrow transplantation (14), detection of changes of tumour versus constitutional DNA fingerprint resulting from cytogenetic

[3]) This paper comprises parts of the dissertation of *Christian Peters*

aberrations and/or DNA rearrangements in tumours (15), and determination of twin zygosity at birth (16).

Here, the simple triplet oligonucleotide, probe, $(GTG)_5$, was hybridized to a panel of human DNAs, which had been digested with the restriction endonucleases *Hinf* I, *Alu* I and *Mbo* I. The resulting DNA fingerprints were analysed with reference to the probability of finding the same banding profile in two unrelated individuals. Subsequently the use of the oligonucleotide probe, $(GTG)_5$, in paternity testing is demonstrated. Examples of paternity and non-paternity are given from actual case work.

Materials and Methods

Genomic DNA of the individuals tested was isolated from peripheral blood following the protocol of *Pöche* et al. (17). DNA (5–10 µg) from each individual was digested with the restriction endonucleases *Hinf* I, *Alu* I and *Mbo* I (GIBCO/BRL, Bethesda USA) according to the supplier's recommendations. The DNA was fractionated on 0.7% horizontal agarose gels in a buffer containing 89 mmol/l Tris-HCl, pH 8.0; 89 mmol/l boric acid; 2 mmol/l EDTA. The oligonucleotide hybridization with the $(GTG)_5$ probe (Fresenius AG, Oberursel, Germany) in the gel were performed as described by *Schäfer* et al. (18).

Results and Discussion

The ^{32}P-labelled oligonucleotide probe, $(GTG)_5$, was hybridized to a panel of *Hinf* I (fig. 1a), *Alu* I (fig. 1b) and *Mbo* I (fig. 1c) digested genomic DNAs of 15 unrelated individuals. In figure 1, about 16 clearly discernible bands in the range of $4-24 \times 10^3$ bases could be visualized per individual, irrespective of the enzyme used. A few signals are found in the same positions in several unrelated individuals, but it is not known and quite unlikely that they stem from the same DNA-locus (19). By analysing the band distributions, estimations on their *mean* allele frequency can be obtained. Calculating the mean allele frequency, two essential aspects have to be taken into account:

i) an unknown probability that additional loci exist in the poorly resolved part of the gel (< 4000 bases) and

ii) the simplifying assumption that co-migrating bands in different individuals are identical alleles of the same locus.

Estimates of mean allele frequency will therefore be maximal values. The actual frequency data probably decrease when a larger panel of indivduals is scored. In general, the mean frequency, q, with which an allele, or band occurs in the population can easily be determined by comparing how often a band of particular size appears in randomly selected DNA profiles. Once the frequency, q, is known, the probability, P, that *any allele* in A is also presented in B, is related to the mean allele frequency, q, of that allele by $P = 2q - q^2$. Thus, assuming a mean allele frequency of $q = 0.136$ in the case of *Mbo* I-digested DNA, the probability, P, of finding a certain band simultaneously in A and B is $P = 2q - q^2 = 0.253$ (tab. 1).

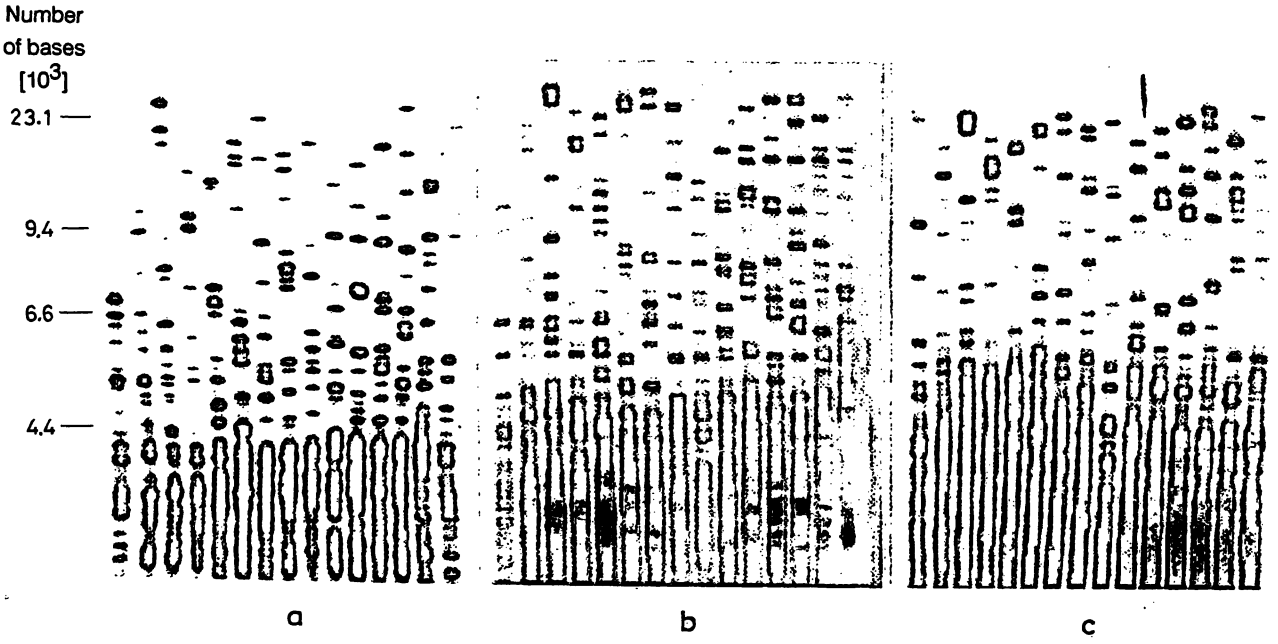

Fig. 1. $(GTG)_5$-DNA fingerprints of 15 unrelated German individuals. DNA (5 µg) from each individual was digested with *Hinf* I (a), *Alu* I (b) and *Mbo* I (c), respectively. Electrophoresis was done in a 0.7% agarose gel at 45 V for 48 h. The gel was hybridized to ^{32}P-labelled $(GTG)_5$ as described in l. c. (18). Note the different hybridization patterns. The sizes of the molecular weight markers are indicated.

Tab. 1. Similarities of $(GTG)_5$-DNA fingerprints between 15 unrelated German individuals (e.g. fig. 1)

Restriction enzyme	*Hinf* I	*Alu* I	*Mbo* I
DNA fragment size (bases)	$4-24 \times 10^3$	$4-24 \times 10^3$	$4-24 \times 10^3$
Number of individuals tested	15	15	15
Total number of bands	243	258	238
Average number (n) of polymorphic bands per individual	16.2	17.2	15.8
Maximal mean allele frequency (q)	0.1895	0.2242	0.1364
Probability (P) of finding an allele in two unrelated individuals $P = 2q - q^2$	0.343	0.398	0.253
Probability (P*) of finding identical banding patterns in two unrelated individuals $P^* = P^n$	2.96×10^{-8}	1.32×10^{-7}	3.7×10^{-10}

DNA samples of the individuals were digested with the three restriction enzymes *Hinf* I (fig. 1a), *Alu* I (fig. 1b), and *Mbo* I (fig. 1c) and hybridized as described in the Materials and Methods section. The bands in each lane were compared with the patterns of the other lanes on the same gel. All clearly distinguishable bands in the range of 4×10^3 bases -24×10^3 bases were scored. The average number (n) of bands per individual was obtained simply by dividing the sum of differing bands by the total number of individuals. Since it is not known which bands represent alleles of the same locus, a mean allele frequency (q) was calculated from the relative frequency of all differing bands according to l.c. (18). The calculations were done assuming that co-migrating bands in different individuals are identical alleles of the same locus. The estimates of allele frequency will thus be maximal. The probability (P) of finding an allele I in two unrelated individuals A and B is then related to the frequency of allele I by $P = 2q - q^2$ and for the whole set of bands by P^n.

Taking an average number of $n = 15.8$ bands per individual into account, the probability, P*, of *identical banding patterns* in two unrelated individuals, A and B, can be calculated as $P^* = P^n = 0.253^{15.8} = 3.7 \times 10^{-10}$.

This implies that theoretically 3×10^9 individuals can clearly be distinguished. Therefore, DNA fingerprinting can be regarded as totally individual-specific except for genetically identical monozygotic twins. The data shown in table 1 corresponds well with the results independently obtained in former examinations (18). The calculated probability P* can even be increased by combining different oligonucleotide probes with several restriction enzymes (e.g. by combining the probes $(GTG)_5$ and $(GACA)_4$ with the restriction endonucleases *Alu* I; *Hinf* I and *Mbo* I).

Because of the high inter-individual discrimination power, DNA fingerprinting is applied to several problems in forensic science e.g. the testing of paternity. For legal fatherhood it can be postulated that the profile of the biological father contains all the bands of the offspring that are not contributed by the maternal genome. Figure 2 shows the $(GTG)_5$ banding patterns of a two-child family obtained with *Hinf* I (a) and *Alu* I (b). Since approximately 50% of the bands in the offspring's profile are contributed by each parent, the degree of shared bands in siblings is in the order of 50%. In this family, no difference in the banding patterns of the two children is exhibited (degree of common bands = 100%), which proves that the two children are monozygous twins. Since all 12 common bands of the children (> 5000 bases) can be traced back to the profile of one of the parents, the putative father can be ascertained as the biological father of the monozygous twins.

To determine the probability for inclusion or exclusion of a putative father from paternity, likelihood ratios, LR, are calculated here according to the method published by *Evett* et al. (20). A likelihood ratio, LR, of e.g. 1000 : 1 means that the likelihood of paternity is 1000-fold larger than the likelihood of non-paternity.

The likelihood ratio, LR, calculated for the family with the monozygous twins reveals a value of LR = 256 085 678 which is equivalent to a paternity probability of W = 99.99%. A W-value of 95% means that given 100 sets (two parents plus child), 95 of the putative fathers will be real fathers and five non-fathers (21). Accordingly, the biostatistical evaluation of this family implies indisputable fatherhood.

Occasionally, an isolated non-matching band might occur in a child due to mutation (9). Therefore the exact mutation rate has to be taken into account when statements on the probability of paternity are given (22). Yet, in figure 3, an example of non-paternity is given: eight bands of the child (> 4000 bases) can neither be referred to the maternal profile nor to the one of the putative father. Since motherhood is not in question, the eight non-relatable bands must be evaluated as an indication of non-paternity.

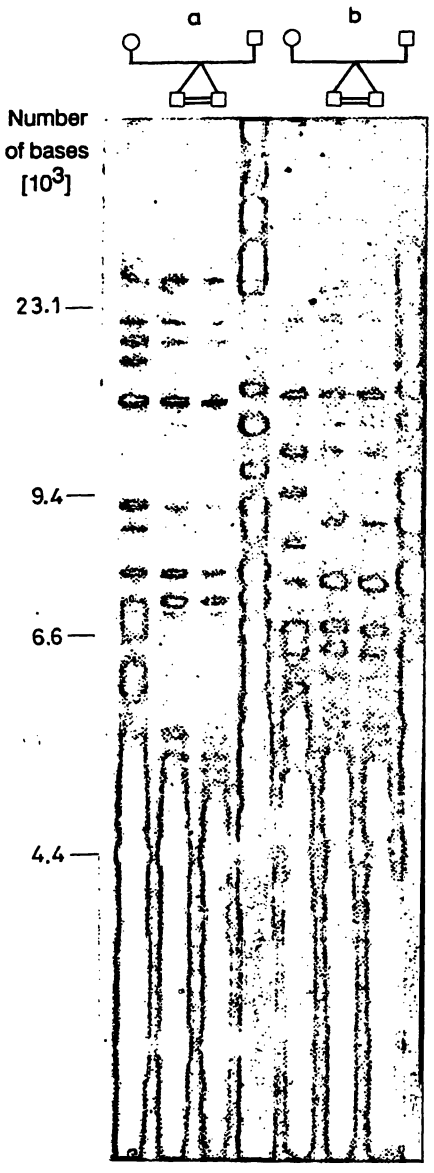

Fig. 2. DNA fingerprints of a family with monozygous twins. DNA (5 µg) from each individual was digested with *Hinf* I (a) and *Alu* I (b). All bands (> 5000 bases) of the twins can be traced back to the parent's profile. A "W"-value of W = 99.99% for paternity indicates biological fatherhood. The sizes of the molecular weight markers are indicated.

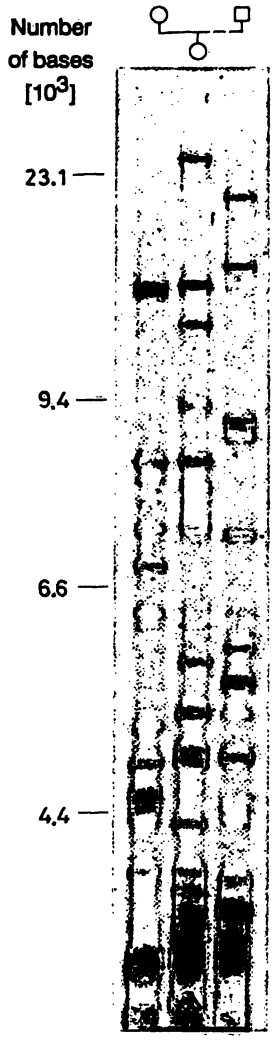

Fig. 3. Paternity testing using $(GTG)_5$ fingerprinting. DNA (5 µg) from each individual was digested with *Hinf* I. Eight bands of the child (> 4000 bases) can neither be found in the maternal profile nor in that of the putative father. Calculating the "paternity probability" a "W"-value of 99.99% *against* paternity excludes the man from fatherhood. The sizes of the molecular weight markers are indicated.

The dual purpose of any paternity test is to identify the falsely accused father, and if not excluded, then to provide a likelihood that this subject is the biological father. The decisive advantage of the DNA fingerprinting technique is that the forensic scientist can come closer than ever before to excluding 100% of the falsely accused fathers, at the same time dramatically increasing the certainity of fatherhood in the case of the non-excluded alleged father.

Finally, a remark on the problem of DNA fingerprinting and data protection should be made: Using *multi*-locus hybridization probes, different, independent loci all over the genome are investigated simultaneously. Furthermore, the fragment length variability of one defined locus overlaps that of many other. Thus, using simple repetitive oligonucleotide and minisatellite probes no information on phenotypic traits, diseases or behavioural patterns of the individuum investigated can be obtained. This implies that data protection is in a way "built into the multilocus system" (23).

Acknowledgement

We thank Mrs. *Gudrun Wrobel* for help and discussion.

References

1. Schmid, C. W. & Shen, C. K. J. (1985) The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. In: *Molecular evolutionary genetics* (MacIntyre, R. J., ed.) Plenum, New York, pp. 323–358.
2. Ridley, R. (1985) Selfish DNA comes of age. New Scientist 1456, 34–37.
3. Cooper, D. N. & Schmidtke, J. (1984) DNA restriction fragment length polymorphisms and heterozygosity in the human genome. Hum. Genet. *66*, 1–16.
4. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. *32*, 314–331.
5. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. & White, R. (1987) Variable number of tandem repeat (VNTR) markers of human gene mapping. Science *235*, 1616–1622.
6. Fowler, C. (1990) Polymorphism in the major human repetitive DNA sequences: theory, practice, and comparison with other methods of measuring human genome variation. In: *DNA in Forensic Science* (Robertson, J., Ross, A. M. & Burgoyne, L. A., eds.) Ellis Horwood, New York, pp. 87–103.
7. Geursen, A. (1990) Hypervariant probes and DNA "fingerprinting". In: *DNA in Forensic Science* (Robertson, J., Ross, A. M. & Burgoyne, L. A., eds.) Ellis Horwood, New York, pp. 74–84.
8. Jeffreys, A. J., Royle, N. J., Wilson, V. & Wong, Z. (1988) Spontaneous mutation rates to new length alleles of tandem-repetitive hypervariable loci in human DNA. Nature *332*, 278–281.
9. Nürnberg, P., Roewer, L., Neitzel, H., Sperling, K., Pöpperl, A., Hundrieser, J., Pöche, H., Epplen, C., Zischler, H. & Epplen, J. T. (1989) DNA fingerprinting with the oligonucleotide probe $(CAC)_5/(GTG)_5$: somatic stability and germline mutations. Hum. Genet. *84*, 75–78.
10. Pöche, H., Wrobel, G., Schneider, V. & Epplen, J. T. (1990) Oligonucleotid-Fingerprinting mit $(GTG)_5$ und $(GACA)_4$ für die Zuordnung von Leichenteilen. Arch. Kriminologie *186*, 37–42.
11. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) Individual-specific "fingerprints" of human DNA. Nature *316*, 76–79.
12. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) Hypervariable "minisatellite" regions in human DNA. Nature *314*, 67–73.
13. Epplen, J. T. (1988) On simple repeated GATA/GACA sequences in animal genomes: a critical reappraisal. J. Hered. *79*, 409–417.
14. Knowlton, R. G., Brown, V., Braman, J. C., Barker, D., Schumm, J. W., Murray, C., Takvorian, T., Ritz, J. & Donis-Keller, H. (1986) Use of highly polymorphic DNA probes for genotyping analysis following bone marrow transplantation. Blood *68*, 378–385.
15. Lagoda, P. J. L., Seitz, G., Epplen, J. T. & Issinger, O. G. (1989) Increased detectability of somatic changes in the DNA from human tumors after probing with "synthetic" and "genome derived" hypervariable multilocus probes. Hum. Genet. *84*, 35–40.
16. Hill, A. V. S. & Jeffreys, A. J. (1985) Use of minisatellite DNA probes for determination of twin zygosity at birth. Lancet *i*, 1394–1395.
17. Pöche, H., Peters, C., Wrobel, G., Schneider, V. & Epplen, J. T. Determining consanguinity by oligonucleotide fingerprinting with $(GTG)_5/(CAC)_5$. Electrophoresis (in press).
18. Schäfer, R., Zischler, H., Birsner, U., Becker, A. & Epplen, J. T. (1988) Optimized oligonucleotide probes for DNA fingerprinting. Electrophoresis *9*, 369–374.
19. Zischler, H., Hinkkanen, A. & Studer, R. (1991) Oligonucleotide fingerprinting with $(CAC)_5$: non-radioactive in gel hybridization and isolation of individual hypervariable loci. Electrophoresis *12*, 141–146.
20. Evett, I. W., Werret, D. J. & Buckleton, J. S. (1989) Paternity calculation from DNA multilocus profiles. J. Forensic Sci. Soc. *29*, 249–254.
21. Hummel, K. (1984) On the theory and practice of Essen-Möller's W-value and Gürtler's paternity index (PI). Forensic Sci. Int. *25*, 1–17.
22. Yassouridis, A. & Epplen, J. T. (1991) On paternity determination from multilocus DNA profiles. Electrophoresis *12*, 221–226.
23. Schacker, U., Schneider, P. M., Holtkamp, B., Bohnke, E., Fimmers, R., Sonneborn, H. H. & Rittner, C. (1990) Isolation of the DNA minisatellite probe MZ 1.3 and its application to DNA fingerprinting analysis. Forensic Sci. Int. *44*, 209–224.

Dr. Hubert Pöche
Institut für Rechtsmedizin
Freie Universität Berlin
Hittorfstraße 18
W-1000 Berlin 33 (Dahlem)
Bundesrepublik Deutschland