The essential point of our report was that early colonies are transient, and that the later-appearing multilineage colonies which contain primitive precursors are not derived from the early colonies. Therefore, the character of later colonies could no longer be invoked as proof of the primitiveness of the cells originating the early colonies. Faced with transient early colonies that had no primitive cells and contained practically only terminally differentiating erythroid cells, and now lacking any evidence for derivation from pluripotential precursors, we proposed the straightforward interpretation: that early and late colonies arise from different kinds of cell, and that the early transient colonies might in fact arise from committed precursors possessing limited proliferative potential.

There is an important additional consequence of the unrelatedness of early and late colonies which concerns the question of probability of 'self renewal'. The concept that this probability could increase during the development of spleen colonies was originally invoked to explain their increasing relative content of primitive precursors with time. An increase in self renewal probability during the course of spleen colony growth is the central assumption in the simulation performed by Blackett *et al.* However, our result removed both the need for this postulate and the evidence supporting it.

The important issue now is whether both early and late spleen colonies arise from an initially homogeneous precursor cell population. There is now abundant evidence that they do not: (1) Eight-day spleen colony-forming units (CFU-S[8]) are killed by 5-fluorouracil, while CFU-S[12] are relatively spared[2]. (2) In sublethally irradiated mice, CFU-S[7] initiate DNA synthesis whereas CFU-S[11] do not[3]. (3) CFU-S[7] are twice as sensitive as CFU-S[11] to photoinactivation by the membrane-binding agent merocyanine 540 (ref. 4). (4) CFU-S[7] are resistant to inactivation by anti-Qa-m2 antibody plus complement, while most CFU-S[12] are inactivated[5]. (5) CFU-S sorted on the basis of binding of pokeweed mitogen segregate into a population high in proliferative capacity and ability to generate primitive precursors, and a second population having lower proliferative capacity and relatively enriched in CFU-S[7] (ref. 6). Sorting on the basis of H33342 staining[7] and anti-H-2K binding similarly segregates cells forming early and late spleen colonies.

These experiments show conclusively that early and late spleen colonies are derived from different and separable types of precursor cells, and directly refute the model proposed by Blackett *et al.* Until contrary evidence appears, the available negative evidence provides more than adequate reason to question the notion that spleen colonies scored at early times arise from pluripotential precursors with extensive proliferative potential.

N. N. ISCOVE
European Molecular Biology Laboratory, Postbox 10 22 09, D-6900 Heidelberg 1, FRG
M. C. MAGLI
Ontario Cancer Institute, 500 Sherbourne Street, Toronto, Canada M4X 1K9
N. ODARTCHENKO
Swiss Cancer Institute, CH-1066 Epalinges sur Lausanne, Switzerland

1. Magli, M. C., Iscove, N. N. & Odartchenko, N. *Nature* **295**, 527–529 (1982).
2. Hodgson, G. S. & Bradley, T. R. *Nature* **281**, 381–382 (1979).
3. Chertkov, J. L. & Drize, N. J. *Cell Tissue Kinet.* **17**, 247–252 (1984).
4. Meagher, R. C. *et al. J. cell. Physiol.* **116**, 118–124 (1983).
5. Harris, R. A. *et al.* Nature **307**, 638–641 (1984).
6. Johnson, G. R. & Nicola, N. A. *J. cell. Physiol.* **118**, 45–52 (1984).
7. Baines, P. & Visser, J. W. M. *Expl Hemat.* **11**, 701–708 (1983).
8. Visser, J. W. M. *et al. J. exp. Med.* **59**, 1576–1590 (1984).

DNA fingerprint analysis in immigration test-cases

RECENTLY, Jeffreys *et al.*[1] used 'fingerprints' of mini-satellites to check maternity in an immigration case. I do not dispute their conclusion that the boy (X) seeking immigration was the son of the putative mother (M), and not the son of an unrelated woman or of a sister of M, but their statistical analysis includes some deficiencies.

A complete analysis of the pedigree, which involved X, M and three undisputed offspring (B, S1 and S2) of M, requires construction of a series of likelihoods taking account of the distribution of DNA bands in each individual, for example information about homozygosity of particular mini-satellites is provided by whether or not all the sibs shared any particular band. It is useful first, however, to reanalyse the summary data given by Jeffreys *et al.*[1] in which, having assumed that X had the same father as B, S1 and S2, they deduced that 25 fragments, all of which were present in M, must have been inherited from X's mother. Following Jeffreys *et al.*, I assume that shared bands represent identical alleles at a hypervariable locus, and that the frequency (x) is the same for all bands. For the model where M is the aunt of X, I have computed the probability that M has the band and X has inherited his from a sister of M (Table 1). The relevant likelihood ratio is $L_a/L_m = 0.63^{25} \doteq 10^{-5}$. Jeffreys *et al.*[1] computed, instead, the probability that M and a sister (assumed, but not stated, not to be an identical twin) share a band, which is not relevant because the band has to be transmitted to X. Their value of 0.62 is so close to 0.63, however, that overall probabilities are little different.

The basic data on the 95 bands observed and the computation of likelihood in terms of the p_{ijk}, the prior probabilities of occurrence of the bands in different family members, are given in Table 2. Computation of p_{ijk} for each model is straightforward, but tedious.

For models unrelated (**u**) and mother (**m**) all possible genotypes of mother and father have to be considered. For example, if both are heterozygotes for a band, the prior frequency of the mating is $4q^2(1-q)^2$, $i=1$ since M has the band, and the conditional probabilities that $k=0$, 1, 2 and 3 are 1/64, 9/64, 27/64 and 27/64, respectively; the conditional probability that $j=0$ is $1/2-q/2$ under model **u** and 1/4 under model **m**. Summation over combinations of genotypes gives the p_{ijk}.

For model **a**, where M is the aunt, a two-stage process can be used. First, all possible matings among the parents of M and her sister and for these the conditional probabilities of each genotype of M and of the band inherited by X from his mother (that is, M's sister) have to be enumerated. For example, if both the grandparents are heterozygotes, the joint probability that M is a heterozygote and X does not inherit the band from his mother is 1/4. From these data, prior probabilities of each of the possible inheritance patterns can be computed as a function of q. Second, these alternatives have to be taken with each possible genotype of the father, and the conditional probabilities that X has the band and the number of sibs with the band computed. For example, if r_3 denotes the prior probability that M is a heterozygote and X does not inherit the band from his mother, the prior probability of the corresponding mating of M to a heterozygote is $2q(1-q)r_3$, $i=1$ (M is a heterozygote), $j=0$ with probability 1/2, and $k=0$ with probability 1/64. Subsequent summation yields the p_{ijk}.

The likelihoods can be obtained for any value of q, but as there is no prior knowledge of band frequency in the Ghanaian population, q has to be estimated for each model by maximizing the likelihood. The important likelihood ratio is $L_a/L_m = 0.000035$ (Table 3) which, though six times higher than the probability ratio calculated by Jeffreys *et al.*, is still very small. If the likelihood computations are repeated with $q=0.14$, the value previously used[1], the corresponding ratio is much reduced (1.3×10^{-7}); and 0.14 was at the upper end of frequencies observed in British caucasians[2]. Also, because frequencies of bands differ[2], these likelihood calculations overestimate L_a/L_m if M is indeed the mother.

There are many assumptions in these analyses, but there seems no justification in avoiding a formal method whatever the

Table 1 Probability that M carries the band and X has a maternally derived band for each model in the reduced analysis

Model	Probability	Likelihood ratio
M unrelated to X (u)	qx	$L_u/L_m = 0.26^{25} = 2.4\times10^{-15}$
M aunt of X (a)	$q(1+x)/2$	$L_u/L_a = 0.413^{25} = 2.5\times10^{-10}$
M mother of X (m)	q	$L_a/L_m = 0.63^{25} = 9.6\times10^{-6}$

x, Band frequency, assumed to be the same for each band ($x=0.26$); q, allele frequency ($x=2q-q^2$).

Table 2 The number of bands (n_{ijk}) present in the sample with $i=(0)1$ if the band is (absent) present in M, $j=(0)1$ if the band is (absent) present in X, $k=0, 1, 2, 3$ is the number of sibs with the band, and computation of likelihood

	Band absent in X ($j=0$)				Band present in X ($j=1$)			
No. of sibs	0	1	2	3	0	1	2	3
Band absent in M ($i=0$)	-	10	4	0	0	7	9	2
Band present in M ($i=1$)	2	5	8	4	3	13	13	15

p_{ijk}, Probability corresponding to n_{ijk}. The log likelihood is computed conditional on the occurrence of the band in at least one individual, with summation excluding the 000 class: $\ln L = \Sigma_i \Sigma_j \Sigma_k n_{ijk}[\ln p_{ijk} - \ln p_{000}]$.

Table 3 Maximum likelihood estimates of allele frequency (q_{max}) for each model, together with corresponding log likelihoods ($\ln L$) and likelihood ratios in the full analysis, assuming B, S1, S2 and X all have the same father

Model	q_{max}	$\ln L$	Likelihood ratios
u	0.32	−256.78	$L_u/L_m = 2.1\times10^{-9}$
a	0.26	−247.05	$L_u/L_a = 5.9\times10^{-5}$
m	0.18	−236.78	$L_a/L_m = 3.5\times10^{-5}$

model. As more information about the frequencies and inheritance of the mini-satellites becomes available, the models can be refined and the obvious power of the DNA-fingerprint technique utilized. Nevertheless, I do not believe that comparisons such as these give positive identification as suggested[1], only evidence that alternatives are highly unlikely.

I thank John Brookfield for providing the basic data, Susan Brotherstone for computational assistance, and members of Leicester University's Genetics Department for financial assistance.

WILLIAM G. HILL
Institute of Animal Genetics
University of Edinburgh,
West Mains Road
Edinburgh EH9 3JN, UK

1. Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* **317**, 818–819 (1985).
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).

JEFFREYS *ET AL.* REPLY—We thank Professor Hill for his rigorous statistical analysis of the DNA fingerprint data of this immigration test-case. We are pleased that the new analysis gives ranges of values for the likelihood ratios which are broadly similar to ours. We reiterate that even the new analysis includes a number of assumptions. The maximum likelihood values of q are based on only one family, and will be inaccurate. It is assumed that there is no variation in q between bands, and it is assumed that bands with the same mobility are always allelic and that bands with different mobilities never are.

We admit that we have not provided positive identification of the mother of X in a mathematical sense, since no probability calculation could do this. However, we believe that the probability of error in this case is so low that it would be legally appropriate to treat the test as providing positive identification.

A. J. JEFFREYS
J. F. Y. BROOKFIELD
R. SEMEONOFF
Department of Genetics,
University of Leicester,
Leicester LE1 7RH, UK

X;autosome translocations in females with Duchenne or Becker muscular dystrophy

RAY *et al.*[1] recently documented their efforts at isolating the gene for Duchenne muscular dystrophy by studying DNA cloned from the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. However, the female patient[2] from whom this X-chromosomal material was derived[3], clearly had Becker muscular dystrophy and not Duchenne dystrophy, as she was 20 years old and still ambulant. It thus seemed timely to critically review all the reported cases of muscular dystrophy in females associated with an X;autosome translocation (Table 1) in order to try and designate them as Duchenne or Becker type.

Duchenne muscular dystrophy is an X-linked condition with onset in early childhood and a rapidly progressive course, so that the vast majority of affected boys lose the ability to walk by the age of 12 years[4]; Becker muscular dystrophy has an identical pattern of muscle involvement but follows a more benign course, and affected males will remain ambulant beyond 16 years of age and usually into adult life[5]. Inevitably there will also be occasional cases who may fall in the grey zone between the two and be a little milder than the usual Duchenne but not quite as mild as the usual Becker type and lose ability to walk between 13 and 16 years.

Careful attention has been paid to the diagnosis of male cases of either Duchenne or Becker dystrophy in the application of restriction fragment length polymorphism (RFLP) studies and detection of deletions. It would seem logical to pay the same attention to the initial diagnosis in the X;autosome translocation females from which the DNA sequences are cloned, rather than loosely referring to them all as Duchenne muscular dystrophy, as seems to have been the current practice.

Of the twelve cases documented to date, four[6–9] conform closely to Duchenne dystrophy and a further four[10–13] also seem to have a Duchenne severity but are still too young to be certain. One additional case[14] is probably a Duchenne but was only published in abstract form with no mention of the age, which had to be surmised from the date of birth and timing of the abstract (approximately 14 years); she was said to be "unable to walk without assistance", which is also difficult to interpret. One case[2] conforms to Becker type and another[15] also seems likely to be, but is too young (13 years) to be certain about ambulation beyond 16 years. The remaining case[16] (published only in abstract form) has insufficient clinical data to draw any conclusions. Two of these cases are somewhat atypical, one[10] having an associated dysmorphic syndrome and the other[9] an associated Turner's syndrome.

Recent studies[17–20] have shown that the gene locus for Becker muscular dystrophy is very close to that for Duchenne and it seems likely that the two conditions are allelic. Studies with some of the probes (for example pERT 87.8)[21] have shown deletions in approximately 9% (5/57) of cases of Duchenne muscular dystrophy but none with cases of Becker muscular dystrophy.

In the report of Ray *et al.*[1], only one Duchenne dystrophy patient out of 50 studied showed a deletion for their XJ-1 clone, and this was also the only case which showed a deletion for pERT 87. In

Received 21 September 1984; accepted 7 January 1985.

1. Bougault, H., Cande, S. C., Schilling, J.-G., Turcotte, D. L. & Olson, P. *Nature* **305**, 278–279 (1983).
2. Richter, F. M. & Ribe, N. M. *Earth planet. Sci. Lett.* **43**, 212–222 (1979).
3. Olson, P., Yuen, D. A. & Balsiger, D. *J. geophys. Res.* **89**, 425–436 (1984).
4. Davies, G. F. *Nature* **290**, 208–213 (1981).
5. Sleep, N. H. *J. geophys. Res.* **89**, 10029–10041 (1984).
6. Morgan, W. J. *Nature* **230**, 42–43 (1971).
7. Schilling, J.-G. *Nature* **242**, 565–571 (1973).
8. Hart, S. R., Schilling, J.-G. & Powell, J. L. *Nature* **246**, 104–107 (1973).
9. Sun, S.-S., Tatsumoto, M. & Schilling, J.-G. *Science* **190**, 143–147 (1975).
10. Schilling, J.-G. & Noe Nygaard, A. *Earth planet. Sci. Lett.* **24**, 1–14 (1974).
11. Morgan, W. J. *J. geophys. Res.* **83**, 5355–5360 (1978).
12. Schilling, J.-G., Kingsley, R. & Devine, J. D. *J. geophys. Res.* **87**, 5593–5610 (1982).
13. Verma, S. P. & Schilling, J.-G. *J. geophys. Res.* **87**, 833–856 (1982).
14. Verma, S.P., Schilling, J.-G. & Waggoner, D. G. *Nature* **306**, 654–657 (1983).
15. Schilling, J.-G., Thompson, G., Kingsley, R. & Humphris, S. *Nature* **313**, 187–191 (1985).
16. Vink, G. E. *J. geophys. Res.* **87**, 9949–9959 (1984).
17. Schilling, J.-G. *et al. Am. J. Sci.* **283**, 510–586 (1983).
18. Neumann, E.-G. & Schilling, J.-G. *Contrib. Miner. Petr.* **85**, 209–223 (1984).
19. Schilling, J.-G. *Trans. Am. geophys. Un.* **58**, 535 (1977).
20. Schilling, J.-G., Vollmer, R. & Bergeron, M. *Trans. Am. Geophys. Un.* **61**, 405 (1980).
21. Morgan, W. J. *The Sea* **7**, 443–487 (1981).
22. Morgan, W. J. *Tectonophysics* **94**, 123–139 (1983).
23. Duncan, R. A. *Tectonophysics* **74**, 29–42 (1981).
24. Duncan, R. A. *J. geophys. Res.* **89**, 9980–9990 (1984).
25. Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331–5354 (1978).
26. Vogt, P. R. *Earth planet. Sci. Lett.* **29**, 309–325 (1976).
27. Anderson, R. N., McKenzie, M. & Sclater, J. G. *Earth planet. Sci. Lett.* **18**, 391–407 (1973).
28. Waggoner, D. G., Vollmer, R. & Schilling, J.-G. *Trans. Am. geophys. Un.* **62**, 423 (1981).
29. Rideout, M. L. & Schilling, J.-G. *IPOD Initial Rep.* **82** (in the press).

Hypervariable 'minisatellite' regions in human DNA

Alec J. Jeffreys*, Victoria Wilson* & Swee Lay Thein†

* Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK
† MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

The human genome contains many dispersed tandem-repetitive 'minisatellite' regions detected via a shared 10–15-base pair 'core' sequence similar to the generalized recombination signal (χ) of Escherichia coli. *Many minisatellites are highly polymorphic due to allelic variation in repeat copy number in the minisatellite. A probe based on a tandem-repeat of the core sequence can detect many highly variable loci simultaneously and can provide an individual-specific DNA 'fingerprint' of general use in human genetic analysis.*

DNA POLYMORPHISMS have revolutionized human genetic analysis and have found general use in antenatal diagnosis[1], mapping of human linkage groups[2,3], indirect localization of genetic disease loci by linkage[2,4,5] and analysis of the role of mitotic nondisjunction and recombination in inherited cancer[6–9]. Single-copy human DNA probes are used normally to detect restriction fragment length polymorphisms (RFLPs), most of which result from small-scale changes in DNA, usually base substitutions, which create or destroy specific restriction endonuclease cleavage sites[10,11]. As the mean heterozygosity of human DNA is low (~0.001 per base pair)[10–12], few if any restriction endonucleases will detect a RFLP at a given locus, although the probability of detection is improved for enzymes such as *Msp*I and *Taq*YI which contain the mutable CpG doublet in their recognition sequence[13]. Even when detected, most RFLPs are only dimorphic (presence or absence of a restriction endonuclease cleavage site) with a heterozygosity, determined by allele frequencies, which can never exceed 50% and which is usually much less. As a result, all such RFLPs will be uninformative in pedigree analysis whenever critical individuals are homozygous.

Genetic analysis in man could be simplified considerably by the availability of probes for hypervariable regions of human DNA showing multiallelic variation and correspondingly high heterozygosities. The first such region was isolated by chance by Wyman and White[14] from a library of random segments of human DNA. The structural basis for multiallelic variation at this locus is not yet known. Subsequently, and again by chance, several other highly variable regions have been discovered near the human insulin gene[15], α-related globin genes[16–18] and the c-Ha-*ras*-1 oncogene[19]. In each case, the variable region consists of tandem repeats of a short sequence (or 'minisatellite') and polymorphism results from allelic differences in the number of repeats, arising presumably by mitotic or meiotic unequal exchanges or by DNA slippage during replication. The resulting minisatellite length variation can be detected using any restriction endonuclease which does not cleave the repeat unit and provides for such loci a set of stably inherited genetic markers.

We have described previously a short minisatellite comprised of four tandem repeats of a 33-base pair (bp) sequence in an intron of the human myoglobin gene (ref. 20; Fig. 1). The 33-bp repeat showed some similarity in sequence to three other hypervariable human minisatellites characterized previously and on the basis that the myoglobin minisatellite was flanked by a 9-bp direct repeat characteristic of the target site duplications generated by transposable elements, we suggested that this minisatellite and some other hypervariable regions were related via transposition. We show here that the myoglobin 33-bp repeat is indeed capable of detecting other human minisatellites, some of which are highly polymorphic. These regions, however, are not related by transposition, but instead share a common short 'core' sequence in each repeat unit, which in turn provides a powerful probe for hypervariable regions.

Probe for variable human DNA

A pure repeat probe was prepared from the human myoglobin minisatellite by purification of a single 33-bp repeat element followed by head-to-tail ligation and cloning of the resulting polymer into pUC 13 (ref. 21; Fig. 1). Cleavage of one of the resulting recombinants, pAV33.7, with *Bam*HI plus *Eco*RI released a 767-bp DNA insert comprised almost entirely of 23 repeats of the 33-bp sequence.

Low stringency hybridization of this repetitive insert to human DNA digested with restriction endonuclease *Eco*RI detected numerous cross-hybridizing DNA fragments, some of which showed signs of polymorphic variation (data not shown). To improve the detection of polymorphisms, the hybridization was repeated with human DNA digested with *Hin*fI or *Hae*III, both of which cleave at a 4-bp sequence not present in the 33-bp repeat sequence and which should release minisatellites in relatively small DNA fragments whose size will reflect more closely the number of repeats per minisatellite. As shown in Fig. 2, the repetitive probe detected multiple DNA fragments in human DNA as well as the parent DNA fragment from the human myoglobin gene. The larger DNA fragments (in the range 2–6 kilobases (kb) and substantially larger than the mean DNA fragment size of ~0.3 kb in human DNA digested with *Hin*fI or *Hae*III) in particular showed variation between the three individuals examined; these variants were transmitted apparently in a Mendelian fashion, in that each polymorphic

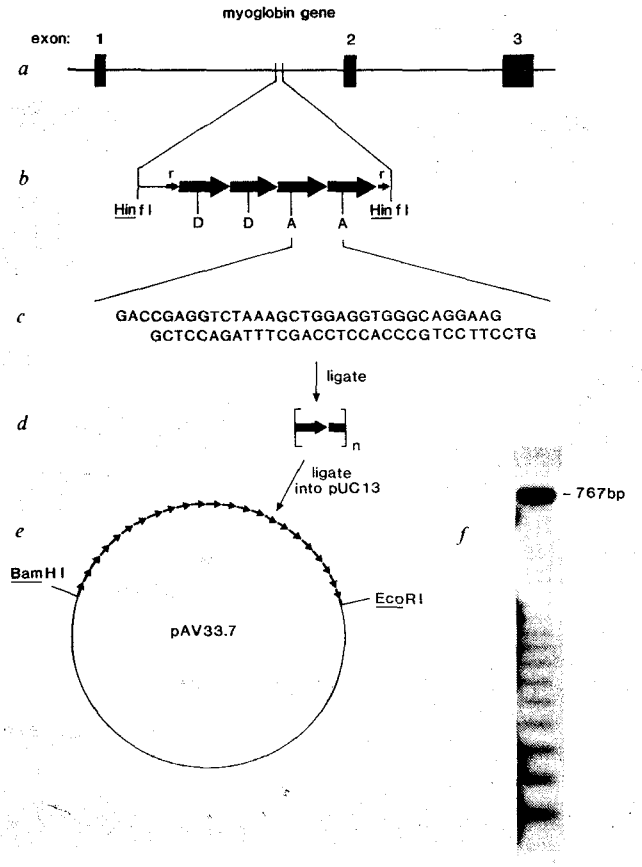


Fig. 1 Construction of a tandem-repetitive hybridization probe for 33-related DNA sequences. *a*, *b*, This probe was derived from a tandem repetitive segment of the human myoglobin gene[20]. This region, located in the first intron and comprising four repeats of a 33-bp sequence flanked by a 9-bp direct repeat (r), was isolated in a 169-bp *Hin*fI fragment, end-repaired and amplified by cloning into the *Sma*I site of pUC13 (ref. 21). *c*, A 33-bp repeat monomer was isolated by cleaving the third and fourth repeat with *Ava*II (A); a single base substitution in repeats 1 and 2 eliminates this site and creates instead a *Dde*I (D) cleavage site[20]. *d*, Ligation of the 33-bp monomer via the non-identical *Ava*II sticky ends produced a head-to-tail polymer. Polymers containing ≥10 repeats were isolated by preparative agarose gel electrophoresis[33], end-repaired, ligated into the *Sma*I site of pUC13 and cloned in *E. coli* JM83 (ref. 21). *e*, The structure of clone pAV33.7 was confirmed by excision of the insert at the polylinker with *Bam*HI plus *Eco*RI, fill-in labelling with α-^{32}P-dCTP at the *Bam*HI site, and partial digestion with *Ava*II. *f*, Labelled partial digest products were resolved by electrophoresis on a 2% agarose gel. pAV33.7 contains 23 repeats of the 33-bp monomer contained in a 767-bp *Bam*HI/*Eco*RI fragment as shown (*e*).

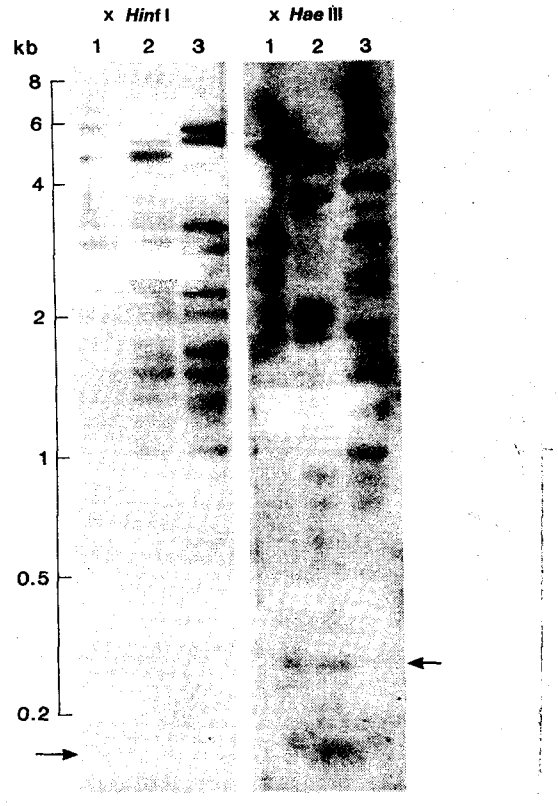


Fig. 2 Detection of multiple 33-related sequences in human DNA. 10 μg samples of DNA from individual 1 (daughter), 2 (mother) and 3 (father) were digested with *Hin*fI or *Hae*III, electrophoresed through a 1% agarose gel and transferred by blotting[35] to a Pall Biodyne membrane to prevent the loss of small DNA fragments. The membrane was hybridized in 1×SSC at 60° with dextran sulphate[36] to the insert from pAV33.7 (Fig. 1) labelled *in vitro* with ^{32}P (ref. 20). The arrowed fragment in each digest is derived from the minisatellite located in the myoglobin gene (Fig. 1); a survey of DNA from 12 individuals digested with *Hae*III showed that this myoglobin minisatellite is monomorphic (data not shown).

band in the daughter could be identified within one or other (but not both) parents. Variation was detectable in both *Hin*fI and *Hae*III digests of human DNA, consistent with polymorphism resulting from length variation of minisatellite regions.

Isolation of minisatellites

A human genomic library[20] of 10–20 kb *Sau*3A partials of human DNA cloned in phage λ L47.1 (ref. 22) was screened by hybridization with the 33-bp repeat probe from pAV33.7. At least 40 strongly-to-weakly hybridizing plaques were identified in a library of 3×10^5 recombinants, consistent with the complexity of the Southern blot hybridization (Fig. 2). A random selection of eight of these positive plaques was purified (λ33.1–15) and Southern blot analysis of phage DNA showed that in each recombinant the hybridizing DNA was localized in a unique short (0.2–2 kb) region of the recombinant. Sequence analysis (Fig. 3) showed that this region in each of the eight recombinants contains a minisatellite of 3–29 tandem copies of a repeat sequence whose length ranged from 16 bp in λ33.15 to 64 bp in λ33.4. Most minisatellites contained an integral number of repeats. In λ33.6, the 37-bp repeat consisted in turn of a diverged trimer of a basic 12-bp unit. Each λ33 recombinant represented a different region of the human genome, judged by the clone-specific DNA sequences flanking each minisatellite.

Highly polymorphic minisatellites

The eight cloned minisatellite regions were located in 0.5–2.2-kb *Hin*fI DNA fragments, smaller than the clearly polymorphic 2–6-kb DNA fragments detected by pAV33.7 in *Hin*fI digests of human DNA (Fig. 2). To determine whether any of the cloned regions were also polymorphic, ^{32}P-labelled single-stranded DNA probes were prepared from suitable M13 subclones of each minisatellite and hybridized at high stringency to a panel of 14 unrelated British caucasian DNAs digested with *Hin*fI. Typical hybridization patterns are shown in Fig. 5, showing that under these hybridization conditions each probe detects a unique region of the human genome. Alleles detected by each probe are summarized in Table 1 where in each case, the most common *Hin*fI allele corresponded in size to the *Hin*fI minisatellite fragment in clones λ33.1–15, suggesting that these regions have been isolated without major rearrangement.

In the limited population sample studied, four of the eight minisatellites showed polymorphic variation. Three of the regions were highly polymorphic with between five and eight resolvable *Hin*fI fragment-length alleles detected per locus. This variation almost certainly results from variation in the repeat

Table 1 Allelic variation at individual cloned minisatellites

Clone λ33	Repeat length (bp)	Divergence %	Alleles			$4N_eu$		
			Length (bp)	No. repeats	Frequency	A	B	C
1	62	0.2±0.2	3,150	40	0.04	2	3	9
			2,600	31	0.11			
			2,350	27	0.04			
			*2,300	26	0.71			
			2,190	24	0.07			
			1,950	20	0.04			
3	32	14.1±2.5	*450	6	1.00	0	0	0
4	64	6.9±1.5	2,280	18	0.07	2	2	5
			2,140	16	0.11			
			*2,015	14	0.43			
			1,950	13	0.36			
			1,780	10	0.04			
5	17	9.2±1.9	*1,660	14	1.00	0	0	0
6	37	0.7±0.4	1,800	25†	0.04	4	6	20
			1,650	21†	0.07			
			1,570	19†	0.04			
			*1,535	18†	0.43			
			1,450	16†	0.04			
			1,400	15†	0.25			
			1,350	14†	0.04			
			1,280	12†	0.11			
10	41	5.9±0.6	*1,460	5	1.00	0	0	0
11	33	0.0±0.0	*990	3	1.00	0	0	0
15	16	1.1±0.5	1,410	41	0.50	0.3	0.3	0.3
			*1,220	29	0.50			

DNA was prepared from white blood cells[10] from 14 unrelated British caucasians (seven male, seven female), digested with *Hin*fI and Southern blot hybridized at very high stringency with probes from subcloned minisatellites as described in Fig. 5*a*, *b*. Cloned alleles whose sequences are shown in Fig. 3 are indicated (*). The divergence of each sequenced minisatellite from a hypothetical $(\text{consensus})_n$ sequence is given as the mean percentage unique substitution divergence (±s.e.m.) to correct for variants which have diffused over more than one repeat. For example, λ33.15 shows 13 variants (10 substitutions and three deletions) but only five distinct variants over 29 repeats of a 16-bp sequence, giving a divergence from a $(\text{consensus})_{29}$ sequence of $5/(29\times16)=1.1\%$. The number of repeats in each allele was estimated from DNA fragment size; this estimate for alleles of λ33.6 is approximate (†) because of the trimeric nature of the λ33.6 repeat unit (Fig. 3). Approximate values of $\theta=4N_eu$, where N_e is the effective population size and u is the rate of production of new length variants per locus per gamete, were estimated for each minisatellite from the number of different alleles (n_a) in the sample of 14 individuals by extensive population computer simulations designed to estimate the number of different, selectively equivalent alleles that could be maintained at steady-state in a population according to three recombinational models: A, random unequal crossing over between two alleles at meiosis, giving a new allele comprised of a random-length 5′ segment of one allele fused to random length 3′ segment of the second allele; B, constrained unequal exchange such that an allele mutates at random to gain or lose between one and three repeats; C, constrained slippage causing the gain or loss of only a single repeat. In each model, it is assumed as a first approximation that u is independent of the number of repeats in allele, as for the loci presented there is at most only a two-fold variation in allele length. Simulations were performed for values of θ from 0.1 to 100, using population sizes N_e from 50 to 500, and were continued for 10 N_e generations, steady state being achieved within $\sim N_e$ generations. Results from model A closely approximated the infinite allele model at $\theta<2$, when the expected number of alleles n_a in a sample of i individuals is given by $n_a=\sum_1^{2i}(\theta/\theta+i-1)$(ref. 31). Model C is the charge state model for which n_a has yet to be solved as a function of θ (ref. 32).

copy number in minisatellites, as alleles generally differed in length by an integral number of repeat units. In addition, longer alleles tended to hybridize with minisatellite probes more strongly than shorter alleles (see individual 1 in Fig. 5*a*), again suggesting that longer alleles contain more repeat sequence.

There is considerable variation in the level of repeat sequence homogeneity in each sequenced minisatellite region (Fig. 3). Some minisatellites (for example, λ33.3 and 5) show substantial repeat divergence, suggesting that these regions are not actively undergoing sequence homogenization by unequal exchange[23]; as expected, these regions show no polymorphic variation in repeat number (Table 1). Instead, the highly polymorphic minisatellites (λ33.1, 4 and 6) all show high repeat copy number together with substantial sequence homogeneity of repeats. In addition, base substitutions in the repeat units of hypervariable minisatellites tend to be present in more than one repeat (see, for example, λ33.15 in Fig. 3) which indicates that these minisatellites are actively and repeatedly engaging in unequal exchange, resulting in the diffusion of novel base substitutions across more than one repeat unit[23].

We used computer simulations to estimate the rate of unequal exchange needed to maintain the number of different (neutral) alleles (n_a) seen in our population sample (Table 1). Although space does not allow a detailed description of the population simulations, we find that for a population of N_e diploid individuals starting with a monomorphic minisatellite containing say 30 repeats, n_a reaches a steady state in $\sim N_e$ generations, the mean value of n_a being determined both by the parameter $\theta=4N_eu$, where u is the rate of production of new length alleles per locus per gamete, and by the model of unequal exchange used in the population simulation. Three models have been investigated: (1) random meiotic unequal exchange between minisatellite alleles; (2) constrained sister chromatid exchange; (3) DNA slippage causing the gain or loss of a single repeat (Table 1, models A, B and C respectively). We favour the constrained exchange model because of the tendency for minisatellite base substitution variants to diffuse to non-adjacent repeats (Fig. 3), together with the tendency of different length alleles of minisatellites to differ from each other by several rather than either one or many repeat units (Table 1). For the highly polymorphic minisatellites, we estimate θ to be 2–6. Given that the effective population size N_e for human populations has been estimated at $\sim10^4$ (ref. 24), this gives values of u, the mutation rate to a new length allele, of $\sim0.5\text{–}1.5\times10^{-4}$ per gamete for a minisatellite ~1 kb long.

This value is higher than the base substitution neutral mutation rate in man, which from studies of non-coding DNA sequence divergence in man and higher primates has been

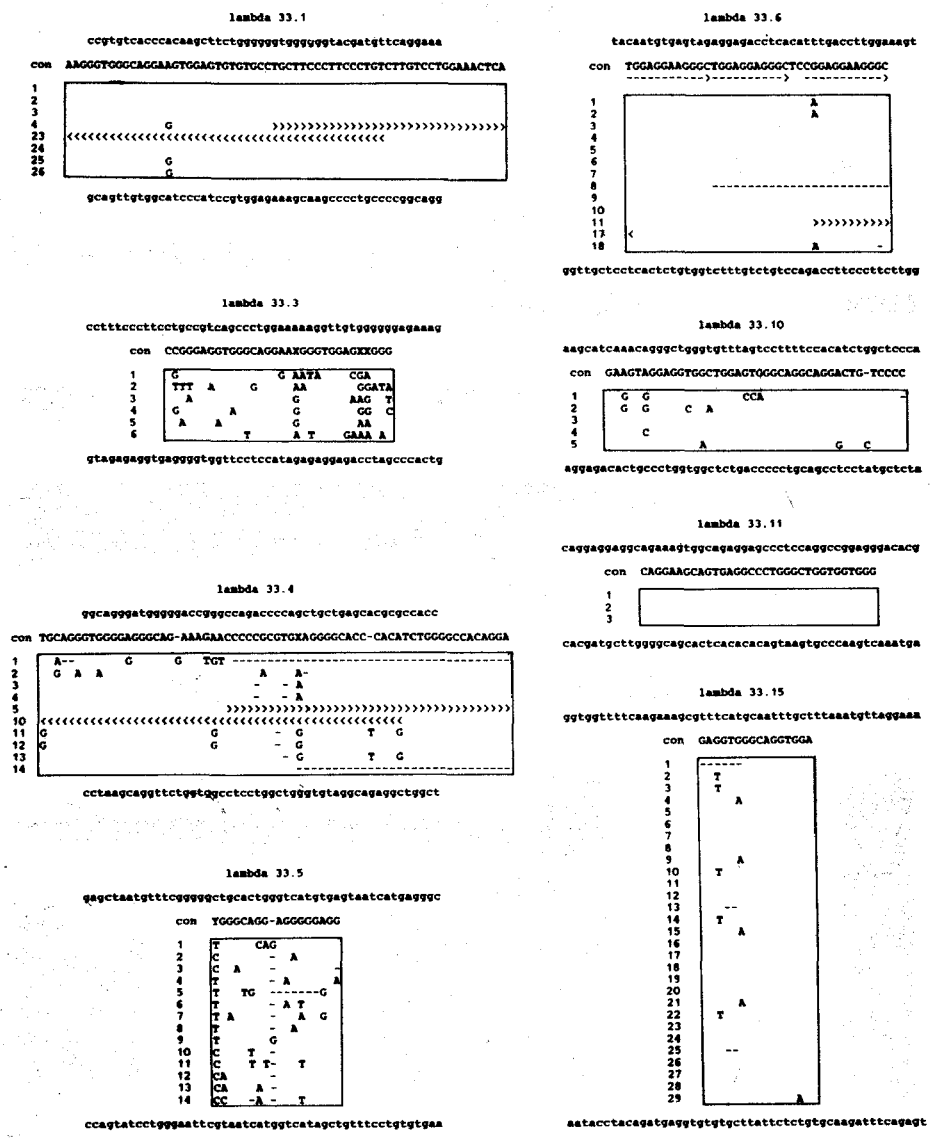


Fig. 3 *Sequences of a selection of* minisatellite regions detected by the myoglobin 33-repeat probe. The consensus sequence (con) of the tandem repetitive region in each of the genomic clones λ33.1–15 is shown, together with 50 bp of 5′ and 3′ flanking DNA (lower case). Differences from the consensus sequences are also shown for the individual numbered repeats (X, A or G; Y, C or T; −, missing nucleotide; »«, *region not sequenced although* clearly a tandem repeat of the consensus sequence, or of a close derivative of the consensus, by inspection of sequencing autoradiographs).

Methods. A library of 10–20-kb human DNA fragments cloned in bacteriophage λL47.1 (refs 20, 22) was screened by hybridization with ^{32}P-labelled pAV33.7 insert as described in Fig. 2. A random selection of eight positive plaques was purified to give recombinants λ33.1–15. Each phage DNA was digested with *Hin*fI or *Hae*III, electrophoresed through a 1.5% agarose gel, and 33-repeat related sequences localized by Southern blot hybridization with pAV33.7 DNA (Fig. 2). Each recombinant gave a single positive *Hin*fI and *Hae*III fragment, except for λ33.4 and 11 which gave no detectable positive *Hae*III fragments (because of the presence of a *Hae*III cleavage site in the repeat regions in these clones; data not shown). Suitable positive *Hin*fI and *Hae*III fragments were isolated by preparative gel electrophoresis[33], end-repaired if necessary and blunt-end ligated into the *Sma*I site of M13mp8 (ref. 37). M13 recombinants were isolated after transformation into *E. coli* JM101 and sequenced by the dideoxynucleotide chain-termination method[38,39]. Each subcloned λ33 fragment contained a tandem repetitive region which in some cases could be sequenced directly. In other cases where the repeat region was too far from the sequencing primer site, the M13 inserts were shortened by cleavage with suitable restriction endonucleases and resequenced.

estimated at 1.0×10^{-9} substitution per nucleotide site per year[25–27]. Assuming that the generation time in man is 20 yr, this predicts a base substitution mutation rate of 2×10^{-5} per 1-kb minisatellite per gamete, lower than the estimated unequal exchange rate of 10^{-4} per gamete. This disparity in rates[23] is probably sufficient to maintain the amount of repeat sequence homogeneity seen in the hypervariable minisatellites λ33.1, 4 and 6.

The rate of unequal exchange can therefore be as high as 10^{-4} per kb minisatellite sequence and presumably is proportional to minisatellite length. In contrast, the rate of homologous recombination at meiosis in human DNA is ~1 centimorgan per 10^6 bp (ref. 2) or 10^{-5} per kb. The apparently very high rate of unequal exchange in minisatellites suggests either that they are hotspots for meiotic recombination, or that most exchanges are between sister chromatids at mitosis in the germline.

A χ sequence in minisatellites?

The length and sequence of the consensus minisatellite repeat sequences vary considerably; none of them are flanked by direct repeats (Fig. 3), in contrast to the repeat region in the myoglobin gene (Fig. 1). Thus, it is unlikely that these minisatellites are related by transposition of a common ancestral sequence. We therefore used dot-matrix comparisons[28] of each minisatellite repeat consensus with the myoglobin 33-bp repeat sequence to determine which region(s) of the 33-bp repeat probe were detecting each minisatellite. Remarkably, the consensus sequence of each minisatellite repeat aligns with the myoglobin repeat specifically over a unique 10–15-bp core region of the 33-bp probe sequence (Fig. 4). This shared core region consists of an almost invariant sequence GGGCAGGAXG preceded by a 5-bp sequence common to most, but not all, repeats.

This core region in each cloned minisatellite suggests strongly that this sequence might help to generate minisatellites by promoting the initial tandem duplication of unique sequence DNA and/or by stimulating the subsequent unequal exchanges required to amplify the duplication into a minisatellite. As polymorphic minisatellites may also be recombination hotspots (see above), it might be significant that the core sequence is similar in length and in G content to the χ sequence, a signal for generalized recombination in *E. coli*[29] (Fig. 4*a*). Although the precise function of χ is unknown, current recombination models[30] suggest that this sequence binds the *rec*BC gene product, endonuclease V, which unwinds locally and nicks DNA to produce a single-stranded DNA projection required for the generation of Holliday junctions. In principle, DNA repair synthesis from the nicking site, followed by ligation to the single-stranded DNA projection, could generate a short tandem duplication with each duplicate containing a χ sequence capable of promoting unequal exchange and amplification of the duplicated region to produce a minisatellite (Fig. 4*b*). Although this model is highly speculative, it predicts that isolated core (or core-like) sequences may also be hotspots for initiating human chromosome recombination.

Probe for hypervariable regions

The repeat length of each minisatellite region is usually half (λ33.5 and 15), the same (λ33.3, 6, 10 and 11) or double the length (λ33.1 and 4) of the 33-bp probe from the human myoglobin gene (Fig. 3). This suggests that detection of minisatellites by pAV33.7 depends not only on the presence of a core sequence in each repeat but also on an in-phase alignment of cross-hybridizable core sequences in a heteroduplex between a minisatellite and the 33-bp repeat probe. If this is correct, then a probe consisting only of a tandem-repeated core sequence should be able to detect not only the human DNA fragments detectable by pAV33.7, but also additional minisatellites incapable of forming stable heteroduplexes with the 33-bp repeat probe. We therefore used the minisatellite from λ33.15, comprising 29 almost identical repeats of an almost perfect 16-bp core sequence, as a hybridization probe for additional minisatellites.

As shown in Fig. 5, this repeated core probe detects a complex profile of hybridizing fragments in human DNA digested with *Hin*fI, including most, but not all, of the bands detected previously with the 33-bp repeat probe from pAV33.7 (comparative data not shown). Only the largest (4–20 kb) *Hin*fI fragments can be resolved fully; these show extreme polymorphism to the extent that the hybridization profile provides an individual-specific DNA 'fingerprint'. Large fragment hyperpolymorphism is to be expected as, if the rate of unequal exchange is proportional to minisatellite length, then long minisatellites (~10 kb long) will have a greater unequal exchange rate ($u \sim 0.001$), raising both the heterozygosity and the number of alleles in a population.

Pedigree analysis

To establish that these large highly polymorphic fragments are stably inherited and segregate in a mendelian fashion, we analysed *Hin*fI digests of DNAs taken from an extensive Asian-Indian pedigree of Gujerati origin, including 54 individuals spanning four generations. The Southern blot/gel electrophoresis time was increased to improve the resolution of these large fragments. Typical examples of part of this pedigree are shown in Fig. 5*d*, *e*, confirming that the polymorphic variation is so great that all individuals, even in a single sibship of a first-cousin marriage (Fig. 5*e*), can be distinguished.

The families in Fig. 5*d*, *e* show that most of the large *Hin*fI fragments are transmitted from each parent to only some of the offspring, establishing that most of these fragments are present in the heterozygous state and that the heterozygosity for these large hypervariable fragments must be approaching 100%. Furthermore, inheritance is Mendelian in that these heterozygous bands are transmitted on average to 50% of the offspring; 48 clearly-resolved different heterozygous parental bands were scored in four sibships of 4–6 individuals and gave a total of 116 cases in which an offspring had inherited a given parental band, compared with 124 cases where the band had not been transmitted (data not shown). Conversely, all fragments in offspring can be traced back to one or other parent, then in turn to their parents (with one exception, see below) and therefore provide a set of stably inherited genetic markers. No band is specifically transmitted from father to son or father to daughter (Fig. 5*d*), eliminating Y and X linkage respectively and implying that these minisatellite fragments are mainly autosomal in origin. Although it is not yet known from where in the set of autosomes these DNA fragments originate, they are not derived from a single localized region of one autosome. Instead, pairs of parental fragments can be identified which segregate independently in the offspring (Fig. 5*d*). Precisely, a pair of bands AB in one parent (and absent from the other) cannot be allelic, nor linked closely in repulsion, if there is at least one AB or − − offspring; the presence of A− or −B recombinant progeny further establishes lack of tight linkage in coupling between A and B. Careful examination of the original autoradiograph of the family shown in Fig. 5*d* reveals, by these criteria, at least 10 resolvable bands in the mother, eight of which are mutually non-allelic and not linked closely. Two other bands may each be an allele of one of the eight unlinked fragments, in that only A− and −B progeny are observed in the limited number of offspring analysed, although such a small sample is insufficient to prove that such pairs of fragments are alleles of a single locus. We conclude that the core probe can give useful information simultaneously on at least several distinct unlinked hypervariable loci.

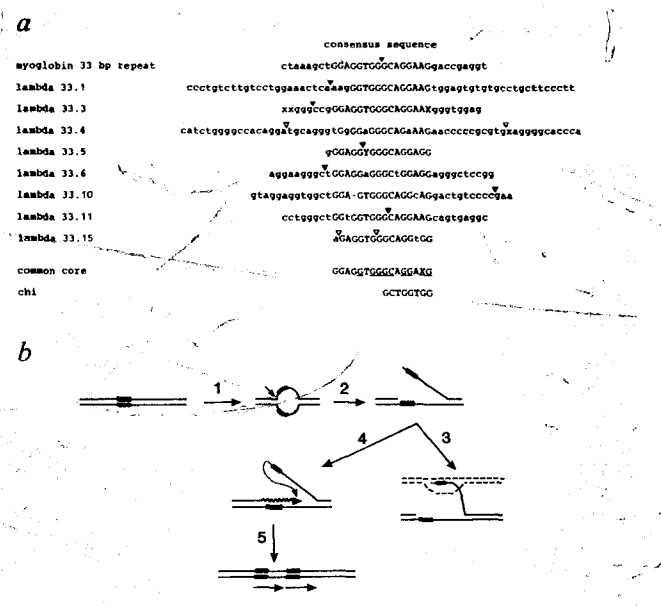


Fig. 4 A common χ-like core sequence shared by the repeat sequence of each minisatellite region. The sequence of each region was compared with the 33-bp tandem repetitive sequence in pAV33.7 and with its reverse complement, using dot-matrix analysis with variable windows and matching criteria[28]. In each case, only a single unambiguous region of sequence similarity was found between the myoglobin 33-bp repeat sequence and the λ33-repeat sequence. The same region was shared by the repeats of all eight λ33 clones. *a*, The aligned sequence of each repeat consensus, each of which is given as an arbitrary cyclic permutation of the consensus shown in Fig. 3. The common core sequence shared by all repeats is also shown and positions in each consensus which conform to the core sequence are identified by upper case letters. Invariant nucleotides in the canonical core sequence are underlined; the generalized recombination signal (χ) of *E. coli* is given also[29]. The beginning/end point of each repeat consensus (Fig. 3) is identified by ▼; in the case of λ33.4 and λ33.15, there is a non-integral number of repeats (Fig. 3) and the separate repeat beginning and end points are shown by ▽. *b*, Model for minisatellite generation promoted by χ-like sequences. (1) A χ-like region denoted by a box. RecBC enzyme binds to χ and unwinds DNA. (2) Nicking produces a single-strand projection (3), which can be assimilated into a homologous duplex to form the precursor of a Holliday junction[30]. (4) Alternatively, DNA repair synthesis followed by (5) ligation and segregation produces a tandem duplication of a χ-containing sequence which can amplify further by unequal exchange. The length of the tandem repeat is determined by the extent of repair synthesis.

A new mutant allele

The extreme variability of the large *Hin*fI DNA fragments detected by the repeat core sequence suggests that the rate of generation of new length alleles must be very high and possibly amenable to direct measurement. There is a clear instance of a new mutation in the pedigree shown in Fig. 5*e*; individual 17 has a new fragment (arrowed), not present in either parent, which might have been derived by unequal exchange and slight expansion of a smaller maternal and paternal fragment present in the other offspring. In a survey of 27 individuals and their parents (data to be presented elsewhere), 240 clearly resolved offspring bands could be traced to one or other parent, the only exception being the band in individual 17. This gives a mutation rate u to a new allele for these hypervariable fragments of

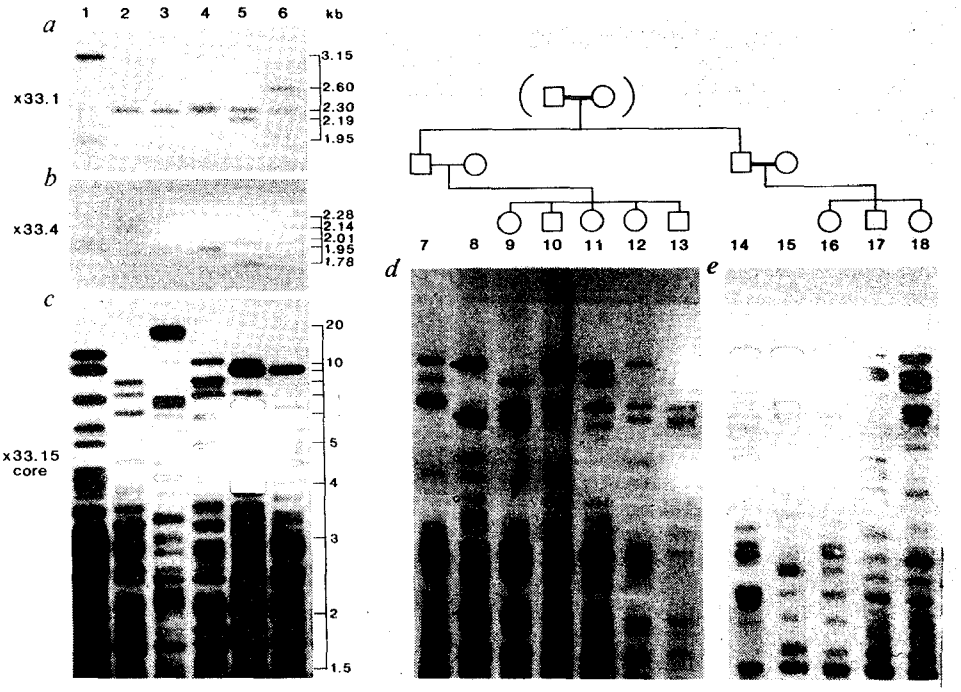


Fig. 5 Polymorphic human DNA fragments detected by hybridization with individual λ33 probes. Southern blots of *Hin*fI digests of DNA from a random sample of British caucasians (1–6) and from selected members of a large British Asian-Indian pedigree (7–18) were hybridized with single-stranded ^{32}P-labelled hybridization probes prepared from suitable M13 recombinants containing minisatellite regions. The pair of bands (○) in individual 8 is an example of non-allelic fragments which are not tightly linked; the pair marked (●) illustrates possible allelism in that each of the five offspring inherits only one of the two fragments. The arrowed fragment in individual 17 is present in neither parent and is a new mutant. The correct paternity of individual 17 has been verified as described below.

Methods. 10 μg samples of DNA prepared from white blood cells[10] were digested with *Hin*fI, electrophoresed through a 20-cm long 1% agarose gel and transferred by blotting to a Sartorius nitrocellulose filter. Single-stranded ^{32}P-labelled hybridization probes were prepared from M13 recombinants as follows. Approximately 0.4 μg M13 single-stranded DNA was annealed with 4 ng 17-mer sequencing primer[40] in 10 μl 10 mM $MgCl_2$, 10 mM Tris-HCl (*p*H 8.0) at 60 °C for 30 min. Primer extension was performed by adding 16 μl 80 μM dATP, 80 μM dGTP 80 μM TTP, 10 mM Tris-HCl (*p*H 8.0), 0.1 mM EDTA plus 3 μl (30 μCi)[α-^{32}P]dCTP (3,000 Ci mmol^{-1}) and 1 μl 5 units μl^{-1} Klenow fragment (Boehringer) and incubating at 37 °C for 15 min. Extension was completed by adding 2.5 μl 0.5 mM dCTP and chasing at 37 °C for a further 15 min. The DNA was cleaved at a suitable restriction endonuclease site either in the insert or in the M13 polylinker distal to the insert, denatured by adding 1/10 vol. 1.5 M NaOH, 0.1 M EDTA, and the ^{32}P-labelled single-stranded DNA fragment extending from the primer was recovered by electrophoresis through a 1.5% low melting point agarose gel (Sea Plaque). The excised band (specific activity > 10^9 c.p.m. μg^{-1} DNA) was melted at 100° C in the presence of 1 mg alkali-sheared carrier human placental DNA (sheared in 0.3 M NaOH, 20 mM EDTA at 100° C for 5 min) and added directly to a pre-warmed hybridization chamber; the carrier DNA also suppressed any subsequent hybridization to repetitive DNA sequences. The precise probes used were: 33.1, a 2,000-nucleotide subcloned *Hae*III fragment containing the minisatellite plus 350-nucleotide flanking human DNA; 33.4, a 695-nucleotide non-minisatellite *Eco*RI fragment on the primer-proximal side of the minisatellite contained in a 2,015-nucleotide *Hin*fI fragment; 33.15 core, a 592-nucleotide subcloned fragment containing the minisatellite sequence plus 128-nucleotide flanking human DNA. Hybridizations were performed as described elsewhere[36], except that dextran sulphate was replaced by 6% (w/v) polyethylene glycol 6,000 (Fisons) to reduce background labelling[41]. Filters A and B were hybridized overnight in 0.5 × SSC at 65 °C and washed in 0.2 × SSC at 65 °C. Filters C–E were hybridized and washed in 1 × SSC at 65 °C. Filters were autoradiographed for 1–3 days at −80 °C using a fast tungstate intensifying screen. The correct paternity of individual 17 was established using a range of biochemical and blood group markers (haptoglobin, transferrin, red cell acid phosphatase, phosphoglucomutase I, adenylate kinase, adenosine deaminase, α_1-antitrypsin, G_c, Gm, esterase D, glyoxylase, phosphoglycollate phosphatase, C3, peptidase D, ABO, Rh and HLA; S.L.T., unpublished data), and confirmed further by rehybridizing this blot with the core minisatellite in λ33.6 to generate a second DNA 'fingerprint' in which all polymorphic bands in individual 17 could be traced back to one or other parent (data not shown).

$\sim\frac{1}{240} = 0.004$. This estimate is in reasonable agreement with the population genetic estimate of $u \sim 0.001$ for these very large hypervariable fragments (see above).

Conclusions

Here, we show not only that it is possible to design probes for the cloning of individual polymorphic minisatellite regions from human DNA, but also that the shared core sequence, which possibly serves as a recombination signal and promotes the formation of minisatellites, can be used for the simultaneous analysis of multiple hypervariable regions. We anticipate that these DNA 'fingerprints' will be of general use in human segregation analysis, in particular for detecting specific bands in close linkage with disease loci in large pedigrees and for studying marker loss in tumours. In addition, they provide a powerful method for paternity and maternity testing, can be used in forensic applications and might also be useful in detecting inbreeding between couples who have had an affected offspring possibly caused by an autosomal recessive gene carried by both parents.

The precise sequence of the core consensus shared by the repeat elements of the cloned minisatellites will be biased by the particular version of the core present in the myoglobin gene minisatellite used as the initial hybridization probe. Therefore, other variant $(core)_n$ probes might detect additional polymorphic loci not found by the λ33.15 repeated core sequence. Preliminary experiments have indeed shown that the core minisatellites in λ33.5 and λ33.6 also hybridize to multiple hypervariable loci, many of which are novel. We are attempting currently to clone large hypervariable regions to provide locus-specific probes for individual minisatellites.

The detection of new mutant-length alleles of minisatellites in human pedigrees will help the analysis of rates and mechanisms of unequal exchange and gene homogenization and can in principle be used to determine whether such exchanges occur by sister chromatid exchange or by recombination between homologous chromosomes at meiosis. In addition, if the core sequence is indeed a recombination signal, then its accurate definition could provide useful substrates for studying mechanisms of human recombination.

This work is the subject of a UK Patent Application. Enquiries should be addressed to National Research Development Corporation. We thank Alain Blanchetot for a human library, John F. Y. Brookfield and Robert Semeonoff for helpful discussions and Raymond Dalgleish for advice on hybridizations. We also thank Drs D. A. Hopkinson, P. Tippett and A. Ting for analysis of various biochemical, blood group and HLA markers. A.J.J. is a Lister Institute Research Fellow; this work was supported by a grant to A.J.J. and a Training Fellowship to S.L.T. from the MRC.

Received 16 November 1984; accepted 14 January 1985.

1. Weatherall, D. J. *The New Genetics and Clinical Practice* (Nuffield Provincial Hospitals Trust, London, 1982).
2. Botstein, D., White, R. L., Skolnick, M. & Davis, R. *Am. J. hum. Genet.* **32**, 314-331 (1980).
3. Drayna, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2836-2839 (1984).
4. Murray, J. M. *et al. Nature* **300**, 69-71 (1982).
5. Gusella, J. F. *et al. Nature* **306**, 234-238 (1983).
6. Cavanee, W. K. *et al. Nature* **305**, 779-784 (1983).
7. Fearson, E. R., Vogelstein, B. & Feinberg, A. P. *Nature* **309**, 176-178 (1984).
8. Koufos, A. *et al. Nature* **309**, 170-172 (1984).
9. Orkin, S. H., Goldman, D. S. & Sallan, S. E. *Nature* **309**, 172-174 (1984).
10. Jeffreys, A. J. *Cell* **18**, 1-10 (1979).
11. Cooper, D. N. & Schmidtke, J. *Hum. Genet.* **66**, 1-16 (1984).
12. Ewens, W. J., Spielman, R. S. & Harris, H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3748-3750 (1981).
13. Barker, D., Schafer, M. & White, R. *Cell* **36**, 131-138 (1984).
14. Wyman, A. & White, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6754-6758 (1980).
15. Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31-35 (1982).
16. Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J.B. & Weatherall, D. J. *Nucleic Acids Res.* **9**, 4213-4224 (1981).
17. Proudfoot, N. J., Gil, A. & Maniatis, T. *Cell* **31**, 553-563 (1982).
18. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5022-5026 (1983).
19. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33-37 (1983).
20. Weller, P., Jeffreys, A. J., Wilson, V. & Blanchetot, A. *EMBO J.* **3**, 439-446 (1984).
21. Vieira, J. & Messing, J. *Gene* **19**, 259-268 (1982).
22. Loenen, W. A. M. & Brammar, W. J. *Gene* **20**, 249-259 (1980).
23. Dover, G. *Nature* **299**, 111-117 (1982).
24. Morton, N. E. *Outline of Genetic Epidemiology*. (Karger, Basel, 1982).
25. Chang, L.-Y. E. & Slightom, J. L. *J. molec. Biol.* (in the press).
26. Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. & Slightom, J. L. *J. molec. Biol.* (in the press).
27. Harris, S., Barrie, P. A., Weiss, M. L. & Jeffreys, A. J. *J. molec. Biol.* (in the press).
28. White, C. T., Hardies, S. C., Hutchison, C. A. & Edgell, M. H. *Nucleic Acids Res.* **12**, 751-766 (1984).
29. Smith, G. R., Kunes, S. M., Schultz, D. W., Taylor, A. & Triman, K. L. *Cell* **24**, 429-436 (1981).
30. Smith, G. R. *Cell* **34**, 709-710 (1983).
31. Ewens, W. J. *Theor. Pop. Biol.* **3**, 87-112 (1972).
32. Ewens, W. J. *Mathematical Population Genetics* (Springer, Berlin, 1979).
33. Dretzen, G., Bellard, M., Sassone-Corri, P. & Chambon, P. *Analyt. Biochem.* **112**, 295-298 (1981).
34. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
35. Southern, E. M. *Meth. Enzym.* **68**, 152-176 (1980).
36. Jeffreys, A. J. *et al. Cell* **21**, 555-564 (1980).
37. Messing, J. & Vieira, J. *Gene* **19**, 269-276 (1982).
38. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
39. Biggin, M. D., Gibson, T. J. & Hong, H. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963-3965 (1983).
40. Duckworth, M. L. *et al. Nucleic Acids Res.* **9**, 1691-1706 (1981).
41. Renz, M. & Kurz, C. *Nucleic Acids Res.* **12**, 3435-3444 (1984).

LETTERS TO NATURE

Cometary impacts, molecular clouds, and the motion of the Sun perpendicular to the galactic plane

Patrick Thaddeus*† & Gary A. Chanan†

* Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA
† Columbia Astrophysics Laboratory, Department of Physics, Columbia University, New York, New York 10027, USA

Raup and Sepkoski[1] have presented evidence from marine fossils for a 26-Myr periodicity in the occurrence of mass extinctions. Using the same data Rampino and Stothers[2] obtained a different period, 30 ± 1 Myr, which agrees with the 33 ± 3-Myr half-period for the vertical oscillation of the Solar System about the plane of the galaxy[3]. To explain this agreement they suggest[2] that encounters with molecular clouds perturb the Sun's family of comets, causing many to enter the inner Solar System where one or more collide with the Earth; the cloud encounter rate is modulated at twice the oscillation frequency, because the number density of clouds peaks at the galactic plane at the midpoint of the solar oscillation crossed by the Solar System twice per period. Notwithstanding an apparent objection to this that the most recent extinctions are not in phase with the solar oscillation[4], their model, given its stochastic nature, can accommodate a few events with large phase discrepancies. The degree of modulation is crucial: it depends on the scale height of the population of molecular clouds relative to the amplitude of the solar motion and tends to zero if this ratio is large and encounters are entirely random. Here we present data from CO surveys of molecular clouds both within and beyond the solar circle, which permit explicit calculation of the strength of the modulation. The cloud layer near the Sun is too extended and, as a consequence, the modulation of cloud encounters is too weak for a statistically significant period to be extracted from the nine extinctions analysed by Rampino and Stothers.

Although we assume a gaussian distribution of clouds with z, the distance from the galactic plane, our conclusions do not depend sensitively on this assumption and remain essentially unchanged if an exponential or other plausible distribution function is assumed instead. The empirical parameter crucial to the mechanism of Rampino and Stothers is, then, the ratio of the amplitude of the solar oscillation z_0 to the half-thickness at half-density $z_{1/2}$ of the distribution of local molecular clouds. One is not free to treat z_0 as a free parameter; it is fixed by (1) the requirement that the period, T, of the oscillation must be 60 Myr, twice the putative extinction period, (2) the circumstance that the Sun is now close to the galactic plane so its z component of velocity, $v_\perp = 7.4\ \mathrm{km\ s^{-1}}$, is essentially that at $z = 0$, and (3) the constraint of simple harmonic oscillation[5], $z_0 = v_\perp T/2\pi = 72$ pc.

Because of the large scatter and uncertainty in the distance of local clouds, there is, as yet, no reliable measurement of $z_{1/2}$ in the vicinity of the sun (that is, within ~1 kpc) from CO or other molecular cloud surveys, but there are good measurements from several large-scale CO surveys of $z_{1/2}$ as a function of galactocentric distance for distant clouds both within the solar circle and beyond. Figure 1 summarizes the survey data. All the CO surveys are consistent with a gradual increase of $z_{1/2}$ with R, proportional approximately to $R^{0.5}$ and a value of $z_{1/2}$ at the solar circle of 85 ± 20 pc (uncertainty 1 s.d.) which we will adopt here for the solar vicinity. Local observations of clouds at known distances are consistent with this value, but are inconsistent with a population of clouds significantly more compressed to the galactic plane. The well-known large concentrations of molecular clouds in Orion and Monoceros[6], for example, representing a significant fraction of local clouds by mass, lie 150–200 pc from the plane, or about twice our adopted $z_{1/2}$.

To calculate the encounter rate as the solar system oscillates sinusoidally through a gaussian distribution of clouds, we may safely neglect the small present displacement of the Sun from the galactic plane. The solar velocity components parallel and perpendicular to the plane are then $v_\parallel = v \cos b = 18.5\ \mathrm{km\ s^{-1}}$ and $v_\perp = v \sin b \sin \omega t = 7.4\ \mathrm{km\ s^{-1}} \sin \omega t$, where $v = 20\ \mathrm{km\ s^{-1}}$ is the present solar motion relative to local stars and interstellar matter[7], $b = 22°$, the galactic latitude of the solar apex[7], and $\omega_0 = 2\pi/60\ \mathrm{Myr^{-1}}$. Because of the low latitude of the apex, the solar speed as a function of time, $v(t) = v_\parallel(1 + 0.164 \cos^2 \omega_0 t)^{1/2} \approx v_\parallel(1 + 0.082 \cos^2 \omega_0 t)$, is always large with respect to the random motion of the clouds, which may be neglected. Letting $\zeta = \ln 2\ z_0^2/z_{1/2}^2$, the number of encounters (extinctions) per unit time is then simply

$$r(t) = n(z(t))v(t)\sigma = C(1 + 0.082 \cos^2 \omega_0 t) \exp(-\zeta \sin^2 \omega_0 t) \quad (1)$$

an explicit function of time at a given ζ with no free parameters except the constant of normalization $C \equiv n_0 v_\parallel \sigma$, where n_0 is the in-plane number density of clouds and σ the encounter cross section. It is tacitly assumed that the encounters are short range, that is, that the impact parameter b is small with respect to $z_{1/2}$; the effect of long range encounters is clearly to reduce the modulation (see below).

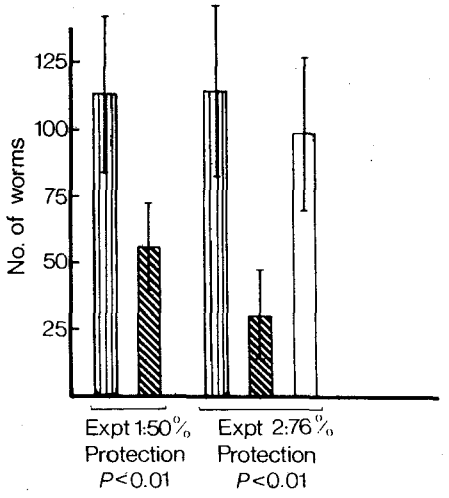


Fig. 4 Protective effect of JM8-36 immunization. LOU rats immunized as described in Fig. 2 legend, giving positive immunofluorescence reactions and significant eosinophil-dependent cytotoxicity, were infected with 800 *S. mansoni* cercariae. Parasite burdens were measured 3 weeks later by liver perfusion[16]. The number of worms obtained from rats immunized with JM8-36 AB_2 antibodies (▨) was compared with those obtained from control groups, that is LOU rats injected with physiological saline (▥) or with IgM purified from normal rat serum (□). The percentage of protection was calculated by the formula $(a-b)/a \times 100$, where a = the number of worms obtained from the saline-injected control group and b = the number of worms recovered from AB_2-immunized rats.

Thus, immunization with anti-idiotype antibodies represents an alternative approach to immunization against pathogens. Although recent studies have produced encouraging results[8-12] concerning the potential substitution of conventional vaccines by anti-idiotype antibodies, this strategy is only in its early stages[13]. In the context of schistosomiasis, idiotype vaccines could be of particular use when relevant protective epitopes cannot easily be produced by the modern tools of molecular biology. Although the rat is a non-permissive host, there is now clear evidence[14] that all the specific effector mechanisms of immunity described in this model also occur in human infection. The possibility of using such immunization procedures in humans remains unexplored and cannot be directly extrapolated from experimental infections. However, work in progress in our laboratory, indicating the existence of cross-reacting idiotypes in human schistosome infection, is encouraging.

We thank Claudine Colson for her help in the preparation of the manuscript. This work was supported by INSERM U167-CNRS 624 and the WHO-UNDP Special Program for Research and Training in Tropical Diseases.

Received 18 February; accepted 24 April 1985.

1. Grzych, J. M., Capron, M., Bazin, H. & Capron, A. *J. Immun.* **129**, 2739-2743 (1982).
2. Dissous, C., Grzych, J. M. & Capron, A. *J. Immun.* **129**, 2232-2234 (1982).
3. Dissous, C., Prata, A. & Capron, A. *J. infect. Dis.* **149**, 227-233 (1984).
4. Grzych, J. M., Capron, M., Dissous, C. & Capron, A. *J. Immun.* **133**, 998-1004 (1984).
5. Jerne, N. K. *Ann. Immun., Inst. Pasteur, Paris* **125C**, 373-389 (1974).
6. Bazin, H., Grzych, J. M., Verwaerde, C. & Capron, A. *Ann. Immun. Inst. Pasteur, Paris* **131D**, 359 (1980).
7. Capron, M. *et al. Eur. J. Immun.* **8**, 127-133 (1978).
8. Sacks, D. L., Esser, K. M. & Sher, A. *J. exp. Med.* **155**, 1108-1119 (1982).
9. McNamara, M. K., Ward, R. E. & Kohler, H. *Science* **226**, 1325-1326 (1984).
10. Stein, K. E. & Söderström, T. *J. exp. Med.* **160**, 1001-1011 (1984).
11. Kennedy, R. C., Melnick, J. L. & Dreesman, G. R. *Science* **223**, 930-931 (1984).
12. Reagan, K. J., Wunner, W. H., Wiktor, T. J. & Koprowski, H. *J. Virol.* **48**, 660-666 (1983).
13. Sacks, D. L. in *New Approaches to Vaccine Development* (eds Bell, R. & Torrigiani, G.) 121-131 (Schwabe, Basel, 1984).
14. Capron, A., Dessaint, J. P., Capron, M. & Joseph, M. in *Progress in Immunology* Vol. 5 (eds Yamamura, Y. & Tada, T.) 1305-1316 (Academic, Tokyo, 1984).
15. Bazin, H., Beckers, A. & Ouerinjean, P. *Eur. J. Immun.* **4**, 44-48 (1974).
16. Smithers, S. R. & Terry, R. J. *Parasitology* **55**, 695-700 (1965).

Individual-specific 'fingerprints' of human DNA

A. J. Jeffreys*, V. Wilson* & S. L. Thein†

* Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK
† MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

Simple tandem-repetitive regions of DNA (or 'minisatellites') which are dispersed in the human genome frequently show substantial length polymorphism arising from unequal exchanges which alter the number of short tandem repeats in a minisatellite[1-4]. We have shown previously that the repeat elements in a subset of human minisatellites share a common 10–15-base-pair (bp) 'core' sequence which might act as a recombination signal in the generation of these hypervariable regions[5]. A hybridization probe consisting of the core repeated in tandem can detect many highly polymorphic minisatellites simultaneously to provide a set of genetic markers of general use in human linkage analysis[5]. We now show that other variant (core)$_n$ probes can detect additional sets of hypervariable minisatellites to produce somatically stable DNA 'fingerprints' which are completely specific to an individual (or to his or her identical twin) and can be applied directly to problems of human identification, including parenthood testing.

Three human minisatellites, termed 33.5, 33.6 and 33.15, each comprised of tandem repeats of various versions of the core sequence, have been cloned previously and characterized[5] (Fig.

Table 1 Similarities of DNA fingerprints between random pairs of individuals

Probe	DNA fragment size (kb)	No. of fragments per individual ±s.d.	Probability x that fragment in A is present in B	Maximum mean allelic frequency/ homozygosity
33.6	10-20	2.8±1.0	0.11	0.06
	6-10	5.1±1.3	0.18	0.09
	4-6	5.9±1.6	0.28	0.14
33.15	10-20	2.9±1.0	0.08	0.04
	6-10	5.1±1.1	0.20	0.10
	4-6	6.7±1.2	0.27	0.14

Samples (8 μg) of blood DNA[6] taken from a random sample of 20 unrelated British caucasians were digested with *Hinf*I and Southern blot hybridized with minisatellite probes 33.6 or 33.15 as described in Fig. 1 legend. Each DNA fingerprint (individual A) was compared with the pattern in the adjacent gel track (individual B), and the number of bands in A which were clearly absent from B, plus those which had a co-migrating counterpart of roughly similar autoradiographic intensity in B, were scored. The data shown are averages for all pairwise comparisons. A small proportion (~6%) of additional weakly hybridizing fragments in A were matched by strongly hybridizing fragments in B, and because in such cases it was not possible to decide whether the band in A was also present in B, such fragments were ignored. If co-migrating bands in A and B are always identical alleles of the same minisatellite locus, then the probability x that an allele in A is also present in B is related to the frequency q of that allele by $x = 2q - q^2$. As the allele frequency is low, then $q^2 \ll q$ and therefore the mean probability $\hat{x}$ is approximately related to the mean allele frequency $\hat{q}$ by $\hat{x} \simeq 2\hat{q}$. Furthermore, assuming that there is little variance in q between alleles, then the number of alleles $n \simeq 1/\hat{q}$ and the mean homozygosity is therefore approximately given by $\sum_1^n q_i^2 \simeq n\hat{q}^2 = \hat{q}$. In practice, an (unknown) proportion of co-migrating bands in A and B will be derived by chance from different minisatellite loci, and thus the estimates of mean allele frequency and homozygosity are maximal and depend on the electrophoretic resolution of minisatellite fragments. Probability estimates: the mean probability that all fragments detected by probe 33.15 in individual A are also present in B is $0.08^{2.9} \times 0.20^{5.1} \times 0.27^{6.7} = 3 \times 10^{-11}$.

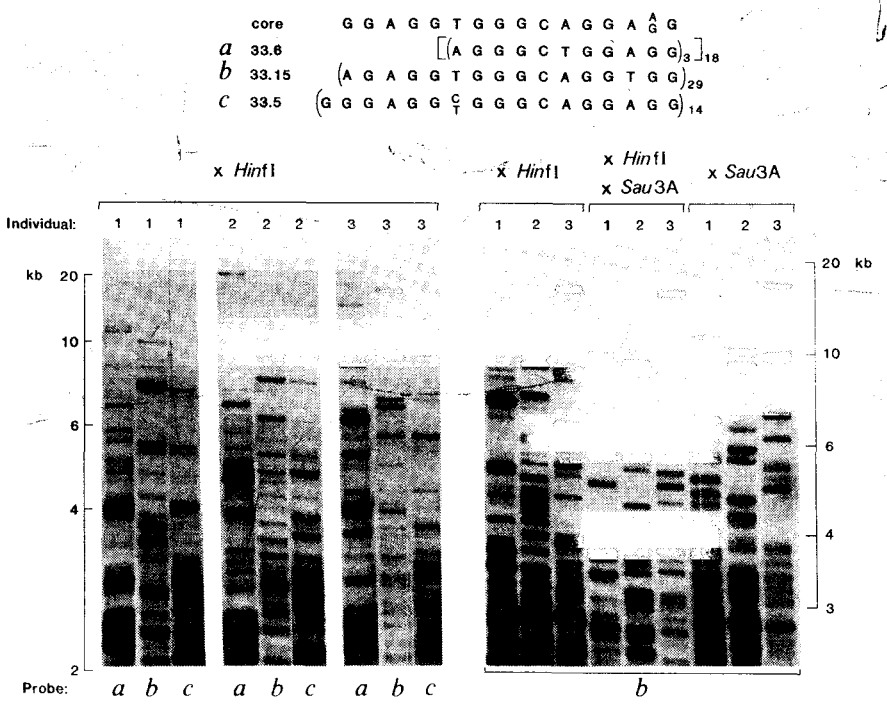


Fig. 1 Hypervariable fingerprints of human DNA. DNA samples prepared from three individual placentae (1-3) were digested with *Hinf*I and/or *Sau*3A and Southern blot hybridized with ^{32}P-labelled single-stranded DNA probes prepared from M13 recombinants 33.5, 33.6 or 33.15, each of which contains a human minisatellite consisting of tandem repeats of closely related variants of the consensus sequences shown; the repeat unit in 33.6 is in turn a diverged trimer[5]. Each probe produces a different fragment pattern whose complexity is largely independent of the tetranucleotide restriction endonuclease used. Resolution of polymorphic fragments less than 4 kb long is improved in double digests with *Hinf*I plus *Sau*3A, due to the elimination of background hybridization caused presumably by relatively diverged and invariant *Hinf*I minisatellite fragments[5] which have accumulated *Sau*3A cleavage sites within one or more repeat units. In double digests, the number of resolvable polymorphic fragments detected by probe 33.15 can be increased from ~15 to ~23 per individual, at the expense of losing ~20% of long single-digest minisatellite fragments which presumably contain a *Sau*3A cleavage site in most or all repeat units.

Methods. DNA was isolated from fresh human placentae as described elsewhere[6]. Samples (8 μg) of DNA were digested with *Hinf*I and/or *Sau*3A, in the presence of 4 mM spermidine trichloride to aid complete digestion, recovered after phenol extraction by ethanol precipitation, and electrophoresed through a 20-cm long 0.6% agarose gel at 30 V for ~24 h, until all DNA fragments <1.5 kb long had electrophoresed off the gel. DNA was then transferred by blotting to a Sartorius nitrocellulose filter[7]. High specific activity (>10^9 c.p.m. ^{32}P per μg DNA) single-stranded M13 probes were prepared as described previously[5]. The precise probes used were: 33.5, a 220-nucleotide *Hae*III fragment containing the minisatellite plus 60 nucleotides of flanking human DNA, subcloned into the *Sma*I site of M13mp8 (ref. 8); 33.6, a 720-nucleotide *Hae*III fragment containing the minisatellite plus 50 nucleotides of flanking human DNA subcloned into the *Sma*I site of M13mp8; 33.15, a 592-nucleotide *Pst*I/*Aha*III fragment containing the minisatellite plus 128 nucleotides of flanking human DNA subcloned into M13mp19 DNA digested with *Pst*I plus *Sma*I. Southern blot hybridization and washing were performed in 1×SSC at 65 °C as described previously[5]. Filters were autoradiographed at room temperature without an intensifier screen for 4 days.

1). Probe 33.15 has been shown to hybridize to multiple hypervariable fragments in *Hinf*I digests of human DNA[5]. Probes 33.5 and 33.6 also detect a complex set of hypervariable regions in human DNA (Fig. 1). Probes 33.5 and 33.15 contain repeats of a similar version of the complete core sequence and consequently produce similar, but not identical, DNA fingerprints. In contrast, probe 33.6 is comprised of a shortened derivative of the core and hybridizes to a largely novel set of hypervariable fragments. Probe 33.15 detects ~15 resolvable hypervariable fragments per individual in the 4–20-kilobase size range, whereas probe 33.6 detects ~11 additional fragments of this size which are not detected by probe 33.15. Probe 33.5 hybridizes on average to about two further large fragments not detected by probes 33.6 or 33.15.

The DNA fingerprint pattern for the longest hypervariable fragments is largely independent of the 4-bp recognition restriction endonuclease used (Fig. 1). This strongly suggests that these large fragments are not derived from longer minisatellites, but that each contains a complete long homogeneous minisatellite devoid of restriction endonuclease cleavage sites and flanked by human DNA containing the normal high density of 4-bp cleavage sites. This is in agreement with previous data showing that most of these large minisatellite fragments are unlinked and segregate independently in pedigrees[5].

To determine the variability of DNA fingerprints, DNA samples from a panel of 20 unrelated British caucasians were digested with *Hinf*I and fingerprinted using probes 33.6 and 33.15 (Table 1). Pairwise comparisons of the DNA fingerprints showed that the minisatellite patterns were highly specific to an individual, and that few fragments are shared between two randomly selected individuals. The probability of shared bands increases for smaller minisatellite fragments, probably resulting from lower genetic variability (higher allele frequencies) of these loci[5] combined with the fortuitous co-migration of unrelated minisatellite fragments. From the degree of band sharing, one can obtain maximal approximate estimates of mean allele frequency and homozygosity (Table 1). For the longest minisatellite fragments in particular, the mean allele frequency is very low (<0.04) and the mean heterozygosity rises to >96%. This is consistent with previous pedigree analysis which has shown that most of these large hypervariable DNA fragments are present in the heterozygous state[5].

The data in Table 1 allow us to estimate the individual specificity of a DNA fingerprint. For probe 33.15, the probability that all 15 resolved fragments in the 4–20-kb size range in an average individual A are also present in a second unrelated individual B is 3×10^{-11} (see Table 1 for details of probability estimates); the probability that the fingerprints of A and B are identical, that is, that all fragments less than 4 kb also match and that B does not possess any additional 4–20-kb fragments, is therefore $\ll 3\times10^{-11}$. Similarly, the probability that A and B have identical fingerprints for both probes 33.15 and 33.6 is $\ll 5\times10^{-19}$. If individuals A and B are related, the chance of fingerprint identity is increased, to a maximum for parents/offspring and sibs. For nonconsanguineous marriages, as the heterozygosity for each fragment is very high, the probability that a hypervariable fragment in sib A is present in sib B is ~1/2, and thus the probability that all 15 resolved bands detected by probe 33.15 in sib A are present in sib B is $\sim 2^{-15} = 3\times10^{-5}$ ($\sim 10^{-8}$ for both probes 33.15 and 33.6). These DNA fingerprints are therefore almost totally individual-specific, even within a single family (except for identical twins, see below).

The DNA fingerprints obtained using these core minisatellite probes are reproducible and are suitable for individual identification. In addition, sufficient DNA (0.5–5 μg) can be isolated rapidly from a single drop of human blood for DNA fingerprinting. This is illustrated in Fig. 2, in which DNA fingerprints were produced from a randomized panel of individuals, including

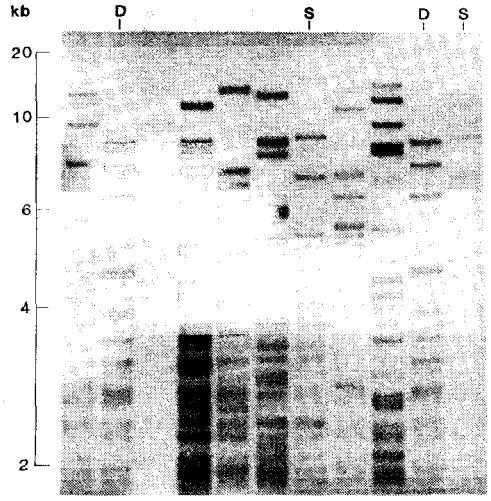


Fig. 2 Individual identification using DNA fingerprints from small samples of blood. DNA was prepared by a rapid procedure using one or two drops of blood from a panel of 11 individuals and digested with *Hinf*I; DNA fingerprints were prepared as described in Fig. 1 legend, using 33.15 as a probe. The panel consisted of nine unrelated individuals, two of whom had been previously DNA fingerprinted, plus two sisters. The autoradiograph was inspected by a colleague who was unaware of the order of samples on the autoradiograph. He correctly identified the two previously fingerprinted individuals on the basis of pattern identity, as well as identifying the two sisters, who have several bands in common (tracks S). He also correctly deduced the fact that duplicate samples had been taken from another individual (tracks D). **Methods.** One or two drops (30–100μl) of blood were collected from a fingerprick into a 1 ml 1×SSC in an Eppendorf centrifuge tube. Cells were pelleted in an Eppendorf microfuge for 1 min, haemolysed by rapid suspension in 1 ml water and immediately made isotonic by the addition of 0.25 ml 5×SSC. White blood cells, nuclei and red cell ghosts were pelleted by centrifugation for 3 min, resuspended in 0.2 ml 0.2 M Na-acetate *p*H 5.6 and lysed by the addition of SDS to 1%. The lysate was extracted twice with phenol/chloroform and DNA collected by two rounds of ethanol precipitation at room temperature. DNA was dissolved in 20 μl water and digested with *Hinf*I in the presence of 4 mM spermidine trichloride at 37°C for 1 h. Electrophoresis and hybridization were performed as described for Fig. 1.

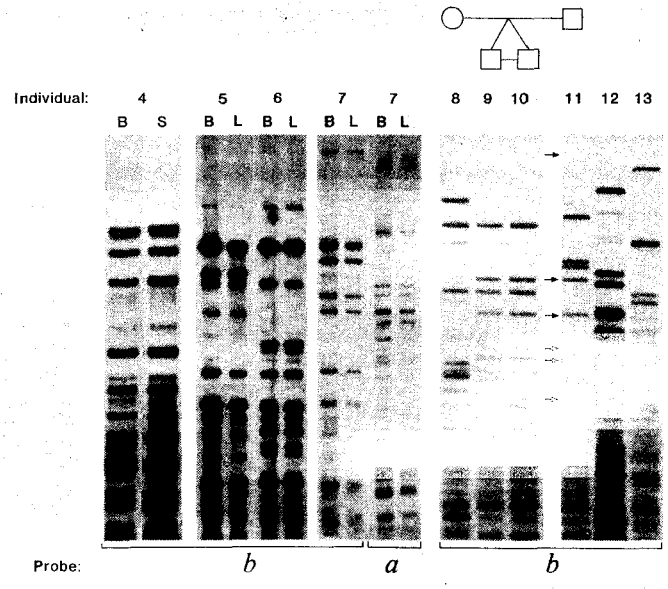


Fig. 3 Somatic stability of DNA fingerprints and their use in paternity testing. Patterns of hypervariable DNA fragments are compared between blood (B) and sperm DNA (S) of individual 4, and between blood DNA (B) and DNA isolated from transformed lymphoblastoid cell lines (L) derived from related individuals 5–7 (6 is the daughter and 7 is the maternal aunt of woman 5; the kinship is evident in the number of fragments shared by these individuals). DNA fingerprints are also shown for blood DNA from identical twins (9, 10) and are compared with their mother (8), father (11) and two unrelated men (12, 13). Resolved paternal DNA fragments in the twins (arrowed) were identified by eliminating maternal bands, and are all present in the father but not in individuals 12 or 13.
Methods. Lymphoblastoid cell lines transformed by Epstein–Barr virus[9] and stored in liquid nitrogen were re-established in liquid culture after 2 yr. These cultured lymphocytes were washed twice in normal saline; DNA from the lymphocyte pellet and from white blood cells was prepared as described elsewhere[6]. Sperm DNA was prepared similarly, except that sperm collected from semen were treated with 1 M 2-mercaptoethanol for 5 min at room temperature before lysis with SDS. DNA fingerprints were prepared as described for Fig. 1, using 5-μg samples of DNA digested with *Hinf*I and hybridized with probe 33.6 (*a*) or 33.15 (*b*).

two people whose DNA had been fingerprinted previously, and two sisters. The two previously characterized individuals could be readily and unambiguously identified on the basis of DNA fingerprint comparisons, as could the two sisters, who have a substantial number of minisatellite fragments in common.

We have shown previously that these hypervariable minisatellite fragments are stably inherited in a mendelian fashion, and that the mutation rate to a new length allele is low (of the order of 0.001–0.004 per locus per gamete for the longest minisatellite fragments)[5]. Several experiments show that these DNA fingerprints are also somatically stable, an essential prerequisite for identification purposes (Fig. 3). Thus, the DNA fingerprints for sperm and blood DNA are indistinguishable, as are the patterns of monozygous twins. Furthermore, the patterns appear to be stably maintained in cultured cells, as shown by comparing the DNA fingerprints of blood DNA with DNA isolated from Epstein–Barr virus-transformed lymphoblastoid cell lines derived from the same individual.

The DNA fingerprints produced by minisatellite probes 33.6 and 33.15 are therefore sufficiently stable and individual-specific for use in human identification in, for example, forensic medicine, and could be used for the routine identification and authentication of human cell lines in culture. They also provide a reliable method for paternity testing (see Fig. 3). Approximately half of the polymorphic minisatellite fragments in an offspring are derived from the father, and these paternal fragments can be identified by comparison of the mother's and offspring's DNA fingerprints. All of these paternal fragments must be present in the father (allowing for a possible rare new mutation). Because, in practice, probe 33.15 will detect approximately six resolved paternal fragments in the 4–20-kb range (Table 1, Fig. 3), we can use the data in Table 1 to estimate the probability that incorrect paternity will not be detected, that is, that the putative father will by chance possess all six paternal-specific DNA fragments. This probability is low ($\sim 5 \times 10^{-5}$) if the putative father is not related to the true father, but increases to $\sim 2^{-6} = \sim 0.016$ if they are closely related (brothers, father/son). If both probes 33.6 and 33.15 are used, these probabilities are reduced to $\sim 4 \times 10^{-8}$ and $\sim 8 \times 10^{-4}$, respectively, although the precise probabilities will depend on the exact resolution and complexity of the DNA fingerprint patterns obtained, and wil be improved if additional paternal fragments <4 kb long can be identified (for example, by using double digests, Fig. 1). We conclude that in the vast majority of cases, the combined use of probes 33.6 and 33.15 will be sufficient to identify cases of incorrect parenthood. An interesting corollary is that these DNA fingerprints could be used with an equal level of reliability to establish true biological parentage.

This work is the subject of a UK patent application. Enquiries should be addressed to the NRDC. We thank John F. Y. Brookfield for helpful discussions, Dr R. S. Pereira for help and advice with lymphoblastoid cell lines, and Dr G. Corney and many

volunteers at Leicester for providing blood samples. A.J.J. is a Lister Institute Research Fellow, and this work was supported by a grant to A.J.J. and a Training Fellowship to S.L.T. from the MRC.

Received 6 March; accepted 3 May 1985.

1. Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31-35 (1982).
2. Proudfoot, N. J., Gil, A. & Maniatis, T. *Cell* **31**, 553-563 (1982).
3. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5022-5026 (1983).
4. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33-37 (1983).
5. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67-73 (1985).
6. Jeffreys, A. J. Cell **18**, 1-10 (1979).
7. Southern, E. M. *Meth. Enzym.* **68**, 152-176 (1980).
8. Vieira, J. & Messing, J. *Gene* **19**, 259-268 (1982).
9. Nilsson, K., Klein, G., Henle, W. & Henle, G. *Int. J. Cancer* **8**, 443-450 (1971).

DNA restriction fragments associated with α_1-antitrypsin indicate a single origin for deficiency allele *PI Z*

Diane Wilson Cox, Savio L. C. Woo* & Tammy Mansfield

Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Canada M5G 1X8 and Departments of Paediatrics, Medical Genetics & Medical Biophysics, University of Toronto, Toronto, M5S 1A8, Canada
* Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA

The α_1-protease inhibitor, or α-antitrypsin (AAT), a major plasma inhibitor of leukocyte elastase and bacterial proteases, is encoded at the *PI* locus on chromosome 14 (14q24.3–q32.1)[1]. A deficiency of AAT in individuals homozygous for the *PI Z* allele occurs in about 1 in 2,000–8,000 caucasians[2] and is associated with an increased risk of early adult onset emphysema[3] and liver disease in childhood[4]. We have now used DNA polymorphisms associated with the AAT gene to investigate the origin of the *PI Z* allele. Using two genomic probes[5] extending into the 5′ and 3′ flanking regions, respectively, we have identified eight polymorphic restriction sites. Extensive linkage disequilibrium occurs throughout the probed region with the *PI Z* allele, but not with normal *PI M* alleles. The *Z* allele occurs mainly with one haplotype, indicating a single, relatively recent, origin in caucasians.

Venous blood samples were collected in EDTA from 32 normal unrelated controls, 26 patients with AAT deficiency (PI type ZZ), and 15 pairs of parents and 16 sibs of patients with AAT deficiency. PI type ZZ patients included 12 adults with emphysema, 3 adults with liver disease, 9 children with liver disease and 2 healthy adults. Relatives of 16 normal controls were studied: members of 2 normal two-generation families with a total of 7 children, and 53 members of a three-generation family of a PI ZZ proband which included 12 unrelated spouses. The genetic type of AAT (PI type) was determined by isoelectric focusing of plasma in acrylamide gels, *p*H 3.5-6 (ref. 6). Blot hybridization of leukocyte DNA digested with specific enzymes is described in Fig. 2 legend. The 5′ probe (4.6 kilobases, kb) and 3′ probe (6.5 kb) have been described elsewhere[7] (Fig. 1), both having been subcloned from phage clone αAT35[5].

Of 10 restriction enzymes tested with the 5′ 4.6-kb probe, 3 showed restriction site polymorphisms; these were *Sst*I, *Msp*I and *Ava*II. For each enzyme, one or both of two possible restriction fragments were observed. When a particular restriction site was present (designated +), a shorter DNA fragment was observed than when the restriction site was absent (designated −). Family studies indicated that the pair of DNA fragments for each enzyme were alleles. Among 32 unrelated normal controls, the allele frequencies were as shown in Table 1. The 4.6-kb genomic probe has been completely sequenced[8], so the positions of the restriction sites were identified. Fragments of the expected sizes were observed. The distances between restriction sites were 2 kb between *Sst*I and *Msp*I sites, and 0.2 kb between *Msp*I and *Ava*II sites.

A χ^2 analysis of 2×2 tables, or exact calculation of probabilities where required[9], was used to examine the data for possible association between a specific allele at one site and a specific allele at each of the other two sites. No significant association between alleles was observed, indicating free recombination between the three restriction sites, with no evidence

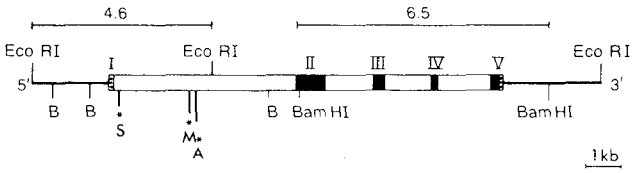


Fig. 1 The α_1-antitrypsin gene and adjacent flanking regions, described previously[4,17]. Solid areas are coding regions, dots indicate untranslated regions, open areas are introns. Restriction sites for the enzymes *Eco*RI and *Bam*HI (B) are indicated. *, Polymorphic restriction sites for *Sst*I (S), *Msp*I (M) and *Ava*II (A). The two genomic probes used in this study are shown above the gene.

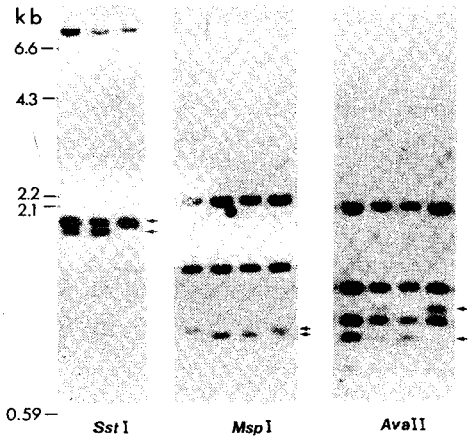


Fig. 2 Patterns of DNA fragments are shown for the three enzymes *Sst*I, *Msp*I and *Ava*II, using the 4.6-kb 5′ probe. DNA extracted from leukocytes was completely digested with specific restriction enzymes according to conditions recommended by the manufacturer. Fragments were separated by size in 0.8% agarose gels. DNA fragments were transferred to Biodyne membranes (Pall). Probes were labelled with ^{32}P using a nick-translation kit (Amersham). After prehybridization, ^{32}P-labelled probe (3.0×10^7 counts per filter) was hybridized overnight. Filters were rinsed twice in $2\times$SSC (SSC: 1.5×10^{-4} M Nacl, 1.5×10^{-5} M Na-citrate). Membranes were washed in $2\times$SSC at 60 °C for 45 min, then twice in 0.1% SDS in $0.1\times$SSC at 60 °C for 45 min, rinsed briefly in SSC, blotted, then exposed to X-ray film for 2-3 days. Fragment sizes in kb are indicated on the left. Anode is at the bottom. Arrows mark the polymorphic fragments. From left to right, lanes show the following allele designations: +−, ++, −−; −−, ++, ++, +−; ++, +−, ++, −−.

Table 1 Allele frequencies for 5′ DNA polymorphisms (4.6-kb probe)

Restriction enzyme	Alleles (kb)	Allele frequency +	Allele frequency −	PIC*
*Sst*I	1.8, 1.9	0.649	0.351	0.35
*Msp*I	0.95, 0.98	0.506	0.494	0.38
*Ava*II	0.9, 1.1	0.636	0.364	0.36

* Polymorphism information content, calculated according to the formula of Botstein *et al.*[21], expressing the extent of variability at each locus

11. Mercer, W. E., Nelson, D., DeLeo, A. B., Old, L. J. & Baserga, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6309–6312 (1982).
12. Jenkins, J. R., Rudge, K. & Currie, G. A. *Nature* **312**, 651–654 (1984).
13. Eliyahu, D., Raz, A., Gruss, P., Givol, D. & Oren, M. *Nature* **312**, 646–649 (1984).
14. Parada, L. F., Land, H., Weinberg, R. A., Wolf, D. & Rotter, V. *Nature* **312**, 649–651 (1984).
15. Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. *Nature* **311**, 33–38 (1984).
16. Newbold, R. F., Overell, R. W. & Connell, J. R. *Nature* **299**, 633–635 (1982).
17. Ruley, H. E. *Nature* **304**, 602–606 (1983).
18. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596–602 (1983).
19. Klein, G. *Cancer Res.* **19**, 343–358 (1959).
20. Newbold, R. F. & Overell, R. W. *Nature* **304**, 648 (1983).
21. Jenkins, J. R., Rudge, K., Redmond, S. & Wade-Evans, A. *Nucleic Acids Res.* **12**, 5609–5626 (1984).
22. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **34**, 861–869 (1981).
23. Coffman, R. L. & Weissman, I. L. *J. exp. Med.* **153**, 269–279 (1981).
24. Laemmli, U.K. *Nature* **277**, 680–685 (1970).

Positive identification of an immigration test-case using human DNA fingerprints

Alec J. Jeffreys, John F. Y. Brookfield & Robert Semeonoff

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

The human genome contains a set of minisatellites, each of which consists of tandem repeats of a DNA segment containing the 'core' sequence, a putative recombination signal in human DNA[1]. Multi-allelic variation in the number of tandem repeats occurs at many of these minisatellite loci. Hybridization probes consisting of tandem repeats of the core sequence detect many hypervariable minisatellites simultaneously in human DNA[1], to produce a DNA fingerprint that is completely individual-specific and shows somatic and germline stability[2]. These DNA fingerprints are derived from a large number of highly informative dispersed autosomal loci and are suitable for linkage analysis in man[3], and for individual identification in, for example, forensic science and paternity testing[2]. They can also be used to resolve immigration disputes arising from lack of proof of family relationships. To illustrate the potential for positive or inclusive identification, we now describe the DNA fingerprint analysis of an immigration case, the resolution of which would have been very difficult and laborious using currently available single-locus genetic markers.

The case concerned a Ghanaian boy born in the United Kingdom who emigrated to Ghana to join his father and subsequently returned alone to the United Kingdom to be reunited with his mother, brother and two sisters. However, there was evidence to suggest that a substitution might have occurred, either for an unrelated boy, or for a son of a sister of the mother; she has several sisters, all of whom live in Ghana. As a result, the returning boy was not granted residence in the United Kingdom. Analysis of conventional genetic markers (ABO, Rh, MN, Se, Pi, Lu, K, Fy, Jk, Gm, Hp, EAP, GLO, PGM, Gc, EsD and HLA; K.Ll. Rogers, personal communication) showed that the woman and boy in dispute were almost certainly related (probability of no relationship = 0.01), but could not determine whether the woman was the boy's mother or aunt. At the request of the family's solicitor, we therefore carried out a DNA fingerprint analysis to determine the maternity of the boy. To complicate matters, neither the father nor any of the mother's sisters was available for analysis. Furthermore, while the mother was certain that the boy was her son, she was not sure about his paternity. DNA fingerprints from blood DNA samples taken from available members of the family (the mother M, brother B, sisters S1 and S2 and the boy X in dispute) were prepared by Southern blot hybridization to two minisatellite probes, each of which detects a different set of hypervariable minisatellites in human DNA[2,3] (Fig. 1).

The first step was to establish the paternity of X from the patterns of hypervariable fragments. Although the father was unavailable, most of his DNA fingerprint could be reconstructed from paternal-specific DNA fragments present in at least one of the three undisputed sibs (B, S1, S2) but absent from M. Of

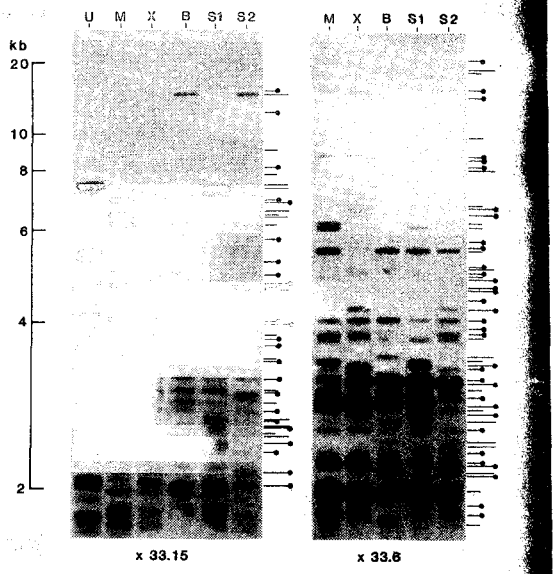


Fig. 1 DNA fingerprints of a Ghanaian family involved in an immigrat dispute. Samples (8 μg) of blood DNA from the mother (M), the boy dispute (X), his brother (B), sisters (S1, S2) and an unrelated individ (U) were digested with *HinfI*, electrophoresed through a 0.7% agarose and Southern blot-hybridized to ^{32}P-labelled minisatellite probe 33.6 or 33 as described elsewhere[1,2]. Fragments present in the mother's (M) DI fingerprints are indicated by a short horizontal line; paternal fragme absent from M but present in at least one of the undisputed sibs (B, S1, S are marked with a long line. 'Maternal' and paternal fragments transmitt to X are shown by a dot. The DNA fingerprints of X contain no additio resolved fragments. All fragments including fainter fragments not visible this photograph were scored from the original autoradiographs taken various exposures; partially resolved fainter bands, particularly towards t bottom of the gel, which could not be reliably scored were ignored. kilobases.

Methods. Quantitation of DNA fingerprints: 61 DNA fragments were scor in M, compared with 39 fragments inherited specifically from the fath One-eighth of the father's heterozygous DNA fragments will not be tra mitted to B, S1 or S2 and thus the corrected estimate for the number paternal-specific fragments is $39 \times 8/7 = 45$. Since the total number of fr ments in the DNA fingerprints of M and the father should be approximat equal, the number of fragments in M which are shared by the father $\sim(61-45) = 16$. The mean probability of band sharing (x) in M and father is therefore $16/61 = 0.26$, consistent with previous estimates der from screening a random sample of Northern Europeans ($x = 0.2$, ref. Approximately half of the 45 paternal bands were transmitted to B, S1 a S2 (18, 24 and 18, respectively), as expected for heterozygous bands. Of t 61 bands in M, more than half were inherited by B, S1 and S2 (32, 38 a 39, respectively; mean = 36.3), as expected since some of M's bands will shared by the father and will therefore be transmitted to most or all childr If M's DNA fingerprints contain n shared bands transmitted to all childr plus $(61-n)$ heterozygous bands transmitted to half the children, then $n+$ $(61-n) = 36.3$, whence $n = 12$, consistent with the estimate of ~16 ba common to M and the father (see above). The DNA fingerprints of X comprised of 21 paternal-specific fragments plus 40 bands shared with The proportion of the latter bands which are maternal-specific and n shared by the father can be estimated in two ways. First, the number *maternal-specific bands should be roughly equal to the number specific* the father, that is, $45/2 = 22.5$. Second, n (~12) of the 40 'maternal' ba in X will be shared maternal/paternal bands (see above), which leaves maternal-specific bands in X. The number of fragments that X has acqui specifically from his mother is therefore ~25. Probabilities of band sharin The mean probability that a fragment in one individual is matched by band of similar electrophoretic mobility and autoradiograph intensity in second random person is defined as x ($x = 0.2$–0.26, see above). Larg minisatellite fragments are less frequently shared, probably due to low allele frequencies and better electrophoretic resolution, and thus the fra ment sharing probability x is heterogeneous[2]. Since almost all fragmen are inherited independently[3], the maximum probability that all n fragmen in an individual are present in a second random individual is therefore x^n any heterogeneity in x will reduce this probability. Band sharing betwee sibs: If shared bands always represent identical alleles of the same hypervar able locus, then, assuming that all alleles have equal frequencies, x is relat to the allele frequency q by $x = 2q - q^2$. At Hardy–Weinberg equilibriu the probability that a band in an individual is also present in a sib can b shown to be $(4+5q-6q^2+q^3)/4(2-q)$ (case i). The other extreme case that bands shared by unrelated individuals are never allelic (that is, th there are many loci at which a band with a given electrophoretic mobili may be found). Then q can be defined as the probability that a given ban will be found in a random gamete from the population. As before, $x = 2q - q$ but the probability of band sharing between sibs can now be shown to $(1+q-q^2)/(2-q)$ (case ii). For $x = 0.26$, q is 0.14 and the probability band sharing is 0.62 (case i) or 0.60 (case ii).

e 39 paternal fragments so identified, approximately half were esent in the DNA fingerprints of X (Fig. 1). Since DNA gments are seldom shared between the DNA fingerprints of related individuals[2] (see individual U in Fig. 1), this suggests ry strongly that X has the same father as B, S1 and S2. After btracting these paternal-specific DNA fragments, there mained 40 fragments in X, all of which were present in M. is in turn provides strong evidence that M is the mother of and therefore that X, B, S1 and S2 are true sibs.

How certain is this identification? We have shown previously at the mean probability that a fragment in the DNA fingerprint one person is present in a second individual selected at ndom is approximately 0.2 for Nothern Europeans[2]. The corre-onding estimate for the father and M is ~0.26 (Fig. 1 legend), ablishing that DNA fingerprint variability in these Ghanaians ot significantly different from that of Northern Europeans. he following probability estimates, we shall make the highly servative assumption that all bands are shared with a uniform bability of 0.26 per band (see Fig. 1 legend for further dis-sion).

he first question is whether X is related to this family. The A fingerprints of X contain 61 scorable fragments, all of ch are present in M and/or the father. If X is unrelated, the bability that each of his bands is present in these parents is $(1-0.26)^2=0.45$; the probability that M and/or the father chance possess all 61 of X's bands is therefore $0.45^{61}=10^{-22}$. X is clearly related to this family.

he next problem is whether an unrelated woman, and not could be the mother of X. The DNA fingerprints of X contain maternal' fragments, of which we estimate that ~25 were rited specifically from the mother; remaining fragments are red between the mother and father and cannot therefore be d to adduce evidence for M's maternity (Fig. 1 legend). All aternal-specific fragments in X are present in M. The chance M is unrelated to X but happens to share all 25 fragments herefore $0.26^{25}=2\times10^{-15}$. Thus, X and M must be related.

he final and most difficult problem is whether any of M's rs, who were not available for analysis, could be the mother (the father, of course, would have to be M's husband). If ds are shared between random people with a mean probabil-f 0.26, the corresponding chance that a fragment in one idual is also present in a sib is 0.62 (Fig. 1 legend). The that M is a sister of X's true mother but by chance contains 5 of X's maternal-specific bands are therefore $0.62^{25}=10^{-6}$. We therefore conclude that, beyond any reasonable t, M must be the true mother of X. This evidence, along results from conventional marker analysis, was provided e immigration authorities, who dropped the case against X granted him residence in the United Kingdom, allowing to remain with his family.

is difficult case demonstrates how DNA fingerprints can unequivocal positive evidence of relationship, even in some where critical family members are missing. The present was simplified by the fact that X has the same father as bs, and that this father did not transmit any bands solely (on average, 1/16th of paternal bands would be so trans-d). Such X-unique fragments, while apparently weakening vidence for the relationship between X and M, would not ractice necessarily invalidate the analysis. X would be ely to have more than 5 such paternal fragments, in addition e 25 maternal-specific fragments. The odds of at least 25 f 30 specified bands matching by chance between X and they are unrelated, or if M is X's aunt, is 8×10^{-11} and 10^{-3}, respectively. This analysis is therefore robust and would clear evidence for or against claimed relationships in most cases. Usually, of course, all relevant members of a family be available, in for example paternity disputes or with lies having difficulties in reuniting by immigration; such will almost always be resolvable using a single DNA rprint test.

thank the family and Sheona York of the Hammersmith ulham Community Law Centre for their kind permission to cite this case, and Dr K. Ll. Rogers for providing blood samples and the results of previous marker tests. A.J.J. is a Lister Institute Research Fellow, and this work was supported by a grant from the MRC. The minisatellite probes are the subject of a Patent Application. Enquiries about the probes and the DNA fingerprint test should be addressed to the Lister Institute of Preventive Medicine, Royal National Orthopaedic Hospital, Brockley Hill, Stanmore, Middlesex, UK.

Received 26 July; accepted 23 September 1985.

1. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–74 (1985).
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).
3. Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. & Ponder, B. A. J. (in preparation).

Repeat sequence families derived from mammalian tRNA genes

Gary R. Daniels & Prescott L. Deininger

Department of Biochemistry, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, Louisiana 70112, USA

Short interspersed repetitive DNA sequences (SINEs) are the major component of dispersed repetitive DNA in all mammalian genomes (see refs 1–6 for reviews). Most SINEs contain an intragenic RNA polymerase III promoter that initiates transcription at the 5′ end of the repeated DNA sequence[7–9] and which has been proposed to facilitate the transposition and amplification of these sequences by an RNA-intermediate mechanism[10,11]. We have discovered several SINE families in the prosimian *Galago crassicaudatus* which have promoter regions similar to transfer RNA genes[12,13]. To determine the relationship between *Galago* SINEs and mammalian tRNA genes, we have compared their sequences. Here, we demonstrate that the *Galago monomer* and type II SINE families[12] are 68 and 62% homologous, respectively, with a human methionine tRNA gene[14]. We have extended our analysis to include the rat identifier[15] and mouse B2 (ref. 16) families and show that their sequences are closely related to alanine and serine tRNA genes, respectively. Our observations suggest that many mammalian SINE families are amplified tRNA pseudogenes.

The most extensively characterized family of SINEs is the human *Alu* family, present in excess of 500,000 copies in the haploid human genome[17]. The *Alu* family consensus sequence corresponds to a dimeric structure composed of two related and tandemly arranged monomeric units, each ~130 base pairs (bp) long with an additional 31 bp in the middle of the right-half sequence (Fig. 1)[18]. This dimeric structure is conserved throughout the primates[13,19], but *Alu* family-related sequences are found only as monomeric sequences in rodent species[1,2,20–22]. This difference in primate and rodent *Alu* family structure suggests that there has been a rapid amplification or replacement of *Alu* family sequences since the rodent–primate divergence.

To determine the mechanisms involved in generating or altering SINEs, we have studied the evolution of *Alu* family structure in several primates and discovered that *G. crassicaudatus* contains two types of *Alu* family sequence (Fig. 1). One is closely related to the human *Alu* family (type I)[13] and the other (type II)[12] has an identical right half, but a very different left-half sequence. We have also discovered that the *Galago* genome contains an independent repetitive DNA family, designated '*monomer*', which is not homologous with the human *Alu* family[12]. The monomer family is similar in copy number to the other two *Galago* families and is closely related to the left-half sequence of the type II family (Fig. 1). The structural relationship of these *Galago* families suggests that the type II sequence has arisen by the integration of a monomer family member into the centre of a type I family[12]. This fusion of two separate repeats has apparently spread as an independent family using the promoter supplied by the monomer sequence

ORF would be predicted to encode a 44-amino-acid hydrophobic protein.

The role of the product of the E5 ORF in the pathogenesis of a fibropapilloma is not known, although it may be involved in the proliferation of the dermal fibroblasts characteristic of a fibropapilloma. ORFs analogous to E5 in BPV-1 have been identified in other papillomavirus genomes that have been sequenced. The sequence of E5 ORFs of HPV-1a (ref. 17) and Shope papillomavirus[18], which induce purely epithelial lesions, however, are not homologous to the E5 ORF of BPV-1. HPV-6b is associated with condyloma accuminata in humans and contains an E5a ORF which shares only slight homology with the E5 ORF of BPV-1[19]. There is, however, a small ORF in the corresponding region of the deer papillomavirus genome[20] and of the BPV-2 genome (W. D. Lancaster, personal communication) which has the potential to encode a peptide strikingly similar in size and amino-acid composition to the BPV-1 E5 ORF. Like BPV-1, the deer papillomavirus and BPV-2 each induces fibropapillomas in their natural host and fibroblastic tumours in hamsters, and readily transforms rodent cells in culture.

The mechanism by which the putative E2 gene product can *trans*-activate the NCR transcriptional regulatory element and its interaction with other viral or cellular genes during the transformation process remains unknown. The dissociation of the *trans*-activation and transformation functions of the 3′ ORFs demonstrated here suggests that the E2 gene product may function indirectly through the activation of the NCR transcriptional regulatory element. The activation of the NCR enhancer by the E2 gene product could lead to increased transcription from the upstream promoter and expression of the 5′ ORFs including E6, E6/E7 and E1, which have been shown to affect transformation and plasmid maintenance[2-4]. However, we have not ruled out the possibility that the E2 ORF gene product may have a direct effect on transformation. Our assay for transformation has been done using the mouse C127 cell line, which is an established cell line. It is possible that the E2 ORF has a direct transformation function but one that must be assayed in a different cell system. Other viral and cellular genes whose products have *trans*-activation properties, including the adenovirus E1a gene[22] and the c-*myc* gene[23], have been shown to interact with other transforming genes in transforming primary cells[24,25]. It is possible that *trans*-activation is only one of several pleiotropic functions of the E2 gene product.

We thank Nan Freas for the preparation of this manuscript. This work was supported in part by a Damon Runyon-Walter Winchell Cancer Fund Fellowship (DRG-028) and an NRSA Fellowship grant (1-F32-CA-07557-01) to B.A.S. and in part by Grant 1597 from the Council for Tobacco Research, USA, Inc. to M.S.R.

Received 5 August; accepted 14 October 1985.

1. Nakabayashi, Y., Chattopadhyay, S. K. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5832-5836 (1983).
2. Sarver, N., Rabson, M. S., Yang, Y.-C., Byrne, J. C. & Howley, P. M. *J. Virol.* **52**, 377-388 (1984).
3. Yang, Y.-C., Okayama, H. & Howley, P. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1030-1034 (1985).
4. Lusky, M. & Botchan, M. R. *J. Virol.* **53**, 955-965 (1985).
5. DiMaio, D., Metherall, J., Neary, K. & Guralski, D. in *Papilloma Viruses: Molecular and Clinical Aspects* (eds Howley, P. & Broker, T.) (Liss, New York, in the press).
6. Spalholz, B. A., Yang, Y.-C. & Howley, P. M. *Cell* **42**, 183-191 (1985).
7. Law, M.-F., Lowy, D. R., Dvoretzky, I. & Howley, P. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2727-2731 (1981).
8. Chen, E. Y., Howley, P. M., Levinson, A. D. & Seeburg, P. H. *Nature* **299**, 529-534 (1982).
9. Heilman, C. A., Engel, L., Lowy, D. R. & Howley, P. M. *Virology* **119**, 22-34 (1982).
10. Amtmann, E. & Sauer, G. *J. Virol.* **43**, 59-66 (1982).
11. Engel, L. W., Heilman, C. A. & Howley, P. M. *J. Virol.* **47**, 516-528 (1983).
12. Stenlund, A., Zabielski, J., Ahola, H., Moreno-Lopez, J. & Pettersson, U. *J. molec. Biol.* **182**, 541-554 (1985).
13. Ahola, H., Stenlund, A., Moreno-Lopez, J. & Pettersson, U. *Nucleic Acids Res.* **11**, 2639-2650 (1983).
14. Rabson, M. S., Yang, Y.-C. & Howley, P. M. *Cancer Cells* Vol. 4 (Cold Spring Harbor Laboratory, New York, in the press).
15. Schiller, J., Vousden, K., Vass, W. C. & Lowy, D. R. *J. Virol.* (in the press).
16. Groff, D. E. & Lancaster, W. D. *Virology* (submitted).
17. Danos, O., Engel, L. W., Chen, E. Y., Yaniv, M. & Howley, P. M. *J. Virol.* **46**, 557-566 (1983).
18. Giri, I., Danos, O. & Yaniv, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1580-1584 (1985).
19. Schwarz, E. *et al. EMBO J.* **2**, 2341-2348 (1983).
20. Groff, D. E. & Lancaster, W. D. *J. Virol.* **56**, 85-91 (1985).
21. Schiller, J. T., Vass, W. C. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7880-7884 (1984).
22. Imperiale, M. J. & Nevins, J. R. *Molec. cell. Biol.* **5**, 875-882 (1984).
23. Kingston, R. E., Baldwin, A. S. Jr & Sharp, P. A. *Nature* **312**, 280-282 (1984).
24. Ruley, H. E. *Nature* **304**, 602-604 (1983).
25. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-602 (1983).
26. Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6453-6457 (1982).
27. Gorman, G. M., Moffat, L. F. & Howard, B. *Molec. cell. Biol.* **2**, 1044-1051 (1982).

Forensic application of DNA 'fingerprints'

Peter Gill*, Alec J. Jeffreys† & David J. Werrett*

* Central Research Establishment, Home Office Forensic Science Service, Aldermaston, Reading, Berkshire RG7 4PN, UK
† Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Many highly polymorphic minisatellite loci can be detected simultaneously in the human genome by hybridization to probes consisting of tandem repeats of the 'core' sequence[1]. The resulting DNA fingerprints produced by Southern blot hybridization are comprised of multiple hypervariable DNA fragments, show somatic and germline stability and are completely specific to an individual[2,3]. We now show that this technique can be used for forensic purposes; DNA of high relative molecular mass (M_r) can be isolated from 4-yr-old bloodstains and semen stains made on cotton cloth and digested to produce DNA fingerprints suitable for individual identification. Further, sperm nuclei can be separated from vaginal cellular debris, obtained from semen-contaminated vaginal swabs, enabling positive identification of the male donor/suspect. It is envisaged that DNA fingerprinting will revolutionize forensic biology particularly with regard to the identification of rape suspects.

Biological samples for forensic analysis consist mainly of bloodstains or semen stains on cloth or other surfaces (often several days or even weeks old), vaginal swabs taken after an alleged rape, and sometimes hair roots[4,5]. Current methods for typing blood and semen stains are reviewed by Divall[6] and Sensabaugh[7]. Typically, a battery of polymorphic protein and blood group marker systems are used (the Metropolitan Police Forensic Science Laboratory, London, currently use ABO, adenylate deaminase (ADA), adenylate kinase (AK), carbonic anhydrase (CA-II), erythrocyte acid phosphatase (EAP), esterase D (EsD), glyoxalase I (GLO), haemoglobin (Hb), peptidase A (Pep A), phosphoglucomutase (PGM_1), Gm, Lewis and rhesus (Rh) for blood analysis). Detection of PGM_1, EAP, ADA and Rh is difficult after 26 weeks[7,8]; EsD and CAII cannot be reliably detected after 1 month[7]. In contrast, only five systems are at present in common use for identifying genetic markers in semen (ABO, PGM, GLO, Pep A and Lewis) and all must be used with considerable caution because semen samples analysed are often contaminated with vaginal fluid which itself has enzyme and blood group activity. Semen contains proteolytic enzymes which further reduce the amounts of proteins recovered[9]. Furthermore, bacterial activity may be responsible for producing erroneous results, particularly in the identification of ABO blood groups[10,11].

DNA fingerprints can only be obtained from high-M_r undegraded DNA. Previous attempts to isolate DNA from dead or aged material have been reported from 140-yr-old quagga muscle[12] and 2,400-yr-old mummies[13]. In both cases only low-M_r DNA was obtained, and it was therefore necessary to establish whether intact DNA could be isolated from samples which may be typically encountered in forensic laboratories.

Multiple samples were prepared from 11 different donors. Bloodstains and semen stains were prepared by aliquoting fresh material onto clean cloth which was air-dried before storage at ambient temperature and humidity. Hair roots were analysed

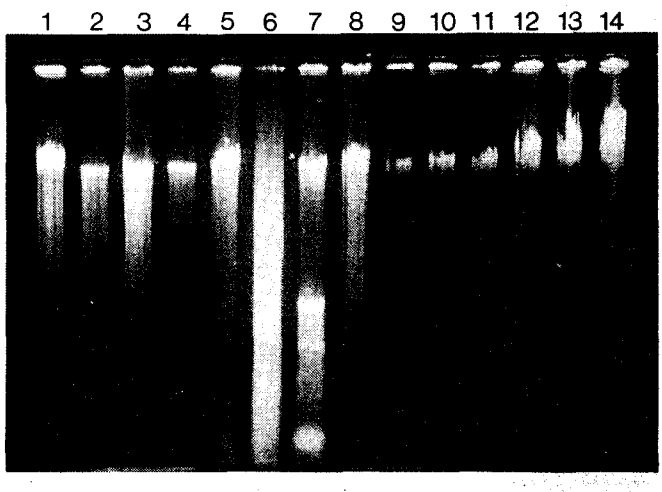


Fig. 1 Isolation of high relative molecular mass DNA from dried bloodstains, semen stains and hair roots. Lane 1, fresh blood (60 μl); lane 2, 6-week-old bloodstain (60 μl equivalent); lane 3, 20-month-old bloodstain (60 μl equivalent); lane 4, fresh semen (20 μl); lane 5, 5-week-old semen stain (20 μl equivalent); lane 6, degraded sample of 5-yr-old semen stain (40 μl equivalent); lane 7, fresh hair roots (×15) (RNA is also clearly visible from hair roots); lane 8, fresh blood (60 μl); lanes 9–14, control human placental DNA, 0.1 μg, 0.2 μg, 0.4 μg, 0.8 μg, 1.6 μg, 3.2 μg respectively.

Methods. Whole blood, semen, hair roots, bloodstains and semen stains were incubated overnight in 1.5-ml microcentrifuge tubes at 37 °C in 0.01 M Tris-HCl, 0.01 M EDTA, 0.1 M NaCl (*p*H 8.0) containing 2% SDS, 20 μg ml^{-1} proteinase K and 0.039 M DTT. DNA was purified by two phenol/chloroform extractions and precipitated by the addition of 0.1 volume of 2 M sodium acetate and 2.5 volumes of absolute ethanol. The DNA was pelleted by centrifugation at 15,000*g* for 5 min, washed with 70% ethanol and repelleted. If a residue remained (suggesting contamination by low-M_r compounds), the sample was dissolved in 20 μl of sterile distilled water and passed down a small spun column (200 μl Sephadex G-50 contained in a 0.75-ml microcentrifuge tube)[15]. Samples were electrophoresed on a 10-cm 0.5% agarose gel at 80 V for 2 h, and DNA was visualized by staining with ethidium bromide.

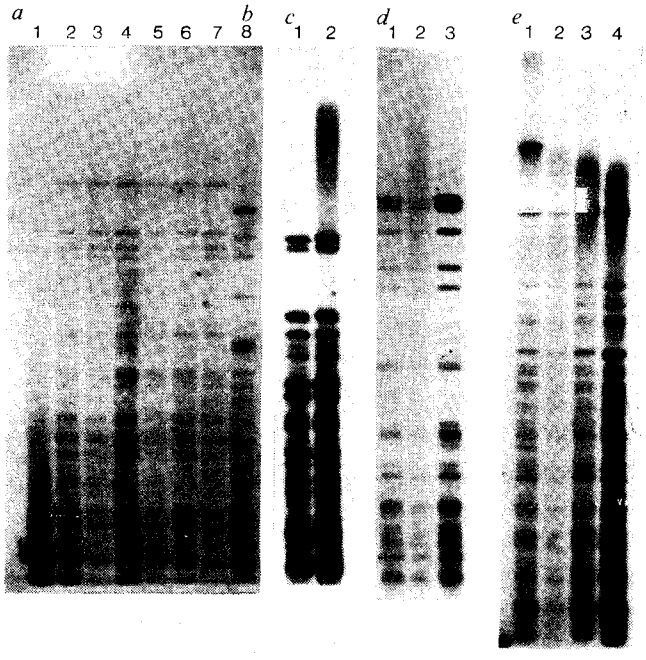


Fig. 2 Somatic stability and ageing studies. Individual *a*: lane 1, degraded sample from 5-yr-old semen stain (40 μl equivalent); lane 2, 4-week-old semen stain (40 μl equivalent); lane 3, fresh semen (40 μl); lane 4, whole blood (60 μl); lane 5, 4-week-old bloodstain (60 μl equivalent); lane 6, 2-yr-old bloodstain; lane 7, fresh hair roots (×15). Individual *b* (showing differences between DNA fingerprints of *a* and an unrelated individual on the same Southern blot): lane 8, fresh blood (60 μl). Individual *c*: lane 1, fresh semen (20 μl); lane 2, 11-week-old semen stain (20 μl equivalent). Individual *d*: lane 1, fresh semen (20 μl); lane 2, 11-week-old semen stain (20 μl equivalent); lane 3, hair roots (×13). Individual *e*: lane 1, fresh blood (60 μl) (the largest band is high-M_r DNA at the electrophoretic exclusion point); lane 2, 1-day-old bloodstain (80 μl equivalent); lane 3, 30-month-old bloodstain (80 μl equivalent); lane 4, 4-year-old bloodstain (80 μl equivalent).

Methods. Samples were prepared as described for Fig. 1. DNA was digested with *Hin*fI (20 units), in the presence of 4 mM spermidine trichloride, for 2 h at 37 °C. Samples were phenol extracted and ethanol precipitated before electrophoresis through a 20-cm long 0.6% agarose gel at 35 V for 20 h. DNA was denatured *in situ*, transferred to a Schleicher and Schull membrane filter (BA85) and filter-hybridized to ^{32}P-labelled single-stranded minisatellite probe 33.15 as described elsewhere[1,2]. Autoradiography was carried out at −80 °C, with an intensifying screen, for 7 days.

fresh. Semen-free and semen-contaminated vaginal swabs (taken at a known period of time after intercourse) were also obtained together with blood samples from both the female and the male. All swabs were stored at −20 °C before use. DNA was extracted from whole blood, whole semen, vaginal fluid, hair roots, bloodstains and semen stains by overnight incubation in an SDS/proteinase K/dithiothreitol (DTT) mixture. Because preliminary experiments showed that semen-contaminated vaginal swabs contained large amounts of DNA from the female, tending to obscure many of the bands from the sperm, female cells were preferentially lysed by preliminary incubation in an SDS/proteinase K mixture. Sperm nuclei are impervious to this treatment because they are ramified with cross-linked thiol-rich proteins[14] and can therefore be separated from the female component by centrifugation. Sperm nuclei were subsequently lysed by treatment with an SDS/proteinase K/DTT mixture. Electrophoresis of samples on 0.5% agarose gels showed that high-M_r DNA could be isolated from bloodstains and semen stains that were at least 2 years old and from fresh hair roots (Fig. 1). However, no DNA could be isolated from hair shafts.

DNA fingerprints obtained from all bloodstains, semen stains and hair roots were shown to be specific to individuals when compared with whole blood and semen samples (Fig. 2). Approximately 5 μl of semen or equivalent semen stain and 60 μl of blood or equivalent bloodstain were required. Some samples analysed were partial digests (Fig. 2*e*) but these also showed patterns consistent with those of a particular individual. DNA fingerprints were obtained from bloodstains up to 4 yr old (Fig. 2*e*). High-M_r DNA from a 5-yr-old semen stain (Fig. 1, lane 6; Fig. 2*a*, lane 1) and a 3-yr-old blood stain had degraded so that no pattern was visualized. After using the differential lysis technique only sperm DNA was isolated from a vaginal swab taken 6.5 h after intercourse (Fig. 3, lane 1).

These preliminary results demonstrate that DNA fingerprints are capable of changing completely the emphasis of blood-grouping in forensic science. Using eight polymorphic protein systems together, Sensabaugh[7] quoted a probability of 0.014 for individual specificity. In practice, the degree of characterization is often much lower, particularly when semen is grouped. This leaves a large degree of uncertainty. In contrast, the condition

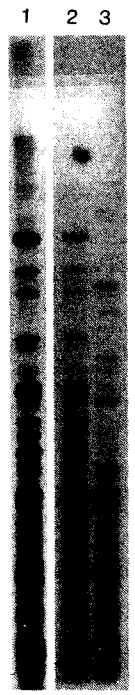


Fig. 3 DNA fingerprint from two vaginal swabs taken 6.5 h after intercourse (lane 1). The pattern is the same as that obtained from the male donor's blood (lane 2). The female's DNA fingerprint (from blood) is shown in lane 3.
Methods. Vaginal cells from semen-contaminated swabs were preferentially lysed by a preliminary incubation in the absence of DTT for 30 min using the lysis mixture described in Fig. 1 legend. Sperm nuclei were pelleted by centrifugation at 15,000*g* for 5 min, washed once, repelleted and lysed with the full SDS/proteinase K/DTT mixture. DNA was isolated as described in Fig. 1 legend and DNA fingerprints prepared (Fig. 2 legend).

of non-association has been absolute using traditional blood-grouping tests, that is, if the phenotype does not match, a common origin is not possible. Using two DNA minisatellite probes (33.15 and 33.6)[1], the degree of association can approach certainty—thus the probability of chance association using probe 33.15 is $<3\times10^{-11}$, and if two probes are used (33.15 and 33.6) the probability is substantially reduced to $\ll5\times10^{-19}$ (ref. 2). In short, using this single method, it is now possible for the forensic scientist to be positive about blood-grouping tests, whereas in the past, it was only possible to be sure of negative associations.

A.J.J. is a Lister Institute Research Fellow and this work was supported in part by a MRC grant to A.J.J. The minisatellite probes are the subject of a Patent Application. Commercial enquiries should be addressed to the Lister Institute of Preventive Medicine, Royal National Orthopaedic Hospital, Brockley Hill, Stanmore, Middlesex HA7 4LP, UK.

Received 7 October; accepted 24 October 1985.

1. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).
3. Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* **317**, 818–819 (1985).
4. Burgess, R. M., Sutton, J. G. & Whitehead, P. H. *J. forens. Sci.* **24**, 392–396 (1979).
5. Gerler, A. & Nagai, T. *Z. Rechtsmed.* **90**, 15–18 (1983).
6. Divall, G. B. *Electrophoresis* **6**, 249–258 (1985).
7. Sensabaugh, G. F. *Isoenzymes: Current Topics in Biological and Medical Research* Vol. 6, 247–282 (Liss, New York, 1982).
8. Denault, G. C., Takimoto, H. H., Kwan, Q. Y. & Pallos, A. *J. forens. Sci.* **25**, 479–498 (1980).
9. Blake, E. T. & Sensabaugh, G. F. *J. forens. Sci.* **21**, 784–796 (1976).
10. Davies, A., Lincoln, P. J. & Martin, P. *Forens. Sci. Int.* **25**, 201–208 (1984).
11. Kishi, K. & Iseki, S. *J. Immunogenet.* **2**, 205–210 (1975).
12. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. *Nature* **312**, 282–284 (1984).
13. Pääbo, S. *Nature* **314**, 644–645 (1985).
14. Calvin, H. I. & Bedford, J. M. *J. Reprod. Fert., Suppl.* **13**, 65–67 (1971).
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Light regulation of plant gene expression by an upstream enhancer-like element

M. P. Timko, A. P. Kausch, C. Castresana, J. Fassler, L. Herrera-Estrella*, G. Van den Broeck*, M. Van Montagu*, J. Schell* & A. R. Cashmore

Laboratory of Cell Biology, Rockefeller University, New York, New York 10021, USA
* Laboratorium voor Genetika, Rijksuniversitiet Gent, Gent B-9000, Belgium

Light regulates many varied physiological and developmental phenomena during plant growth and differentiation, including the formation of a photosynthetically competent chloroplast from a proplastid[1]. The expression of ribulose 1,5-bisphosphate carboxylase small subunit (rbcS) genes is regulated by light in a development- and tissue-specific manner[2,3]. In some plant species, phytochrome has been demonstrated to mediate this response[4–7], and photoregulation of rbcS expression occurs at least in part at the level of transcription[8,9]. We have shown previously that a 5′-noncoding fragment (4–973 base pairs (bp) upstream of the messenger RNA cap site) of the pea rbcS *ss3.6* gene contains all of the nucleotide sequence information necessary to direct the photoregulated expression of a bacterial chloramphenicol acetyltransferase (*cat*) gene in tobacco[10]. Consistent with these findings, Morelli *et al.*[11] have shown by deletion analysis of a second rbcS gene promoter, that the sequences required for photoregulated expression of rbcS E9 reside within the 5′-noncoding region. They identified an upstream region of ~700 bp needed for maximum transcription but not light–dark regulation, and a region from −35 to −2 bp which included the TATA box and contained the necessary information for light responsiveness. We now demonstrate that regulatory sequences 5′ distal to the rbcS *ss3.6* TATA box and transcriptional start site not only contain the information necessary for maximum expression, but also confer photoregulation. These upstream regulatory sequences function independently of orientation when fused to their homologous promoter or a heterologous promoter.

The expression of many animal viral and cellular genes is influenced by regulatory sequences, commonly referred to as enhancers[12–18], which characteristically function independently of orientation at various distances from the TATA box and start site of transcription[12,16–18]. Certain enhancers activate gene expression only in specific cell types[16–18], or in response to physiological stimuli[15]. It has not previously been investigated whether the expression of some plant genes is also regulated by enhancers.

The nucleotide sequence requirement for photoregulated expression of the rbcS *ss3.6* gene[19] was examined using the CAT expression system[20]. We and others[10,20–23] have demonstrated the ability of this sytem to compare promoter efficiency and have shown that for chimaeric genes such as those used here, in which the mRNAs are identical, CAT activity reflects the cellular *cat* mRNA levels[10] and therefore the transcriptional activity directed by the targeted promoter. To overcome any bias due to different RNA processing or translation efficiencies, the various test promoter fusions were always compared with the wild-type promoter.

To define more accurately the sequences within the −973 to −4 rbcS *ss3.6* promoter which are necessary to direct photoregulated expression, a series of promoter deletion mutants was constructed and tested for their ability to direct photoregulated *cat* gene expression in transformed tobacco calli (Fig. 1). Three of the constructs contain 5′ deletions with endpoints at 722, 357 and 92 bp upstream from the cap site and the fourth contains a 3′ deletion lacking sequences −4 to −90 (including the TATA box).

DNA fingerprinting dispute laid to rest

Eric S. Lander and Bruce Budowle

Two principals in the once-raging debate over forensic DNA typing conclude that the scientific issues have all been resolved.

THE US public, usually indifferent to matters scientific, has suddenly become obsessed with DNA. Nightly newscasts routinely refer to the polymerase chain reaction (PCR) and even the tabloids offer commentary on restriction fragment length polymorphisms (RFLPs). The new-found fascination with nucleic acids does not stem from recent breakthroughs in genetic screening for breast cancer susceptibility or progress in gene therapy — developments which will indeed affect the lives of millions. Rather, it focuses on the murder case against the former US football star, O. J. Simpson.

The Los Angeles trial, starting in November and to be broadcast live by several major television networks, will probably feature the most detailed course in molecular genetics ever taught to the US people. This bold experiment in public education should, in principle, be a cause for rejoicing among scientists. The catch is that the syllabus is being prepared by attorneys whose primary roles are as adversaries; the likely result is confusion. Already, the news weeklies are preparing the ground with warnings that DNA fingerprinting remains "controversial", being plagued by major unresolved scientific issues.

Forensic DNA typing certainly did provoke controversy soon after it was introduced into US courts in 1988. The technology itself represents perhaps the greatest advance in forensic science since the development of ordinary fingerprints in 1892, and is soundly rooted in molecular biology. The problem, however, stemmed from the manner of its introduction. Pioneered by biotech start-up companies with good intentions but no track record in forensic science, DNA typing was marred by several early cases involving poorly defined procedures and interpretation[1]. Standards were lacking for such crucial issues as: declaring a match between patterns; interpreting artefacts on gels; choosing probes; assembling databases; and computing genotype frequencies. There is broad agreement today that many of these early practices were unacceptable, and some indefensible. For its part, the US Federal Bureau of Investigation (FBI) moved much more deliberately in developing procedures, sought public comment and opted for conservative procedures.

As a result of these growing pains, forensic DNA typing was subjected to intense debate and scrutiny. When it first burst on the scene, the supporting scientific literature consisted of a mere handful of papers. By the middle of this year, there had been more than 400 scientific papers, 100 scientific conferences, 3 sets of guidelines from the Technical Working Group on DNA Analysis Methods (TWGDAM), 150 court decisions and, importantly, a 3-year study by a National Research Council (NRC) committee released in 1992 (ref. 2). In the light of this extraordinary scrutiny, it seems appropriate to ask whether there remains any important unresolved issue about DNA typing, or whether it is time to declare the great DNA fingerprinting controversy over.

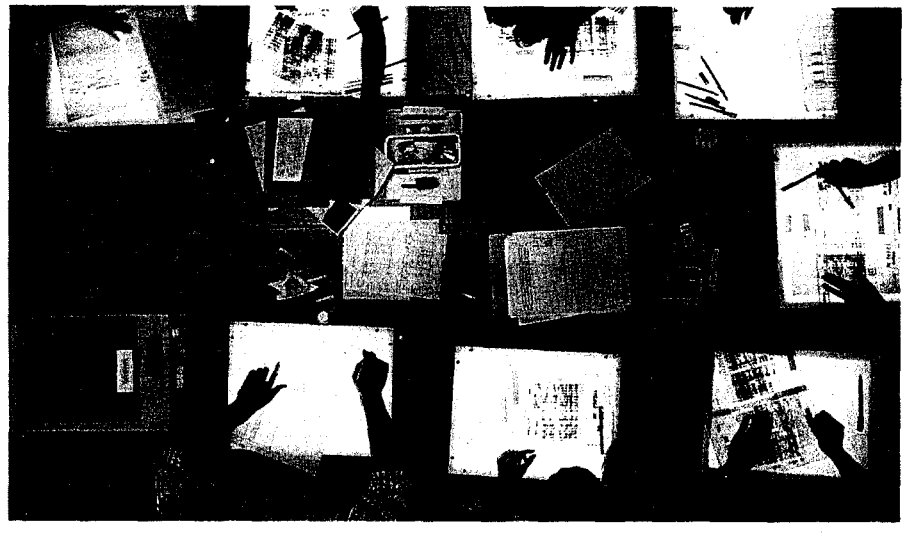
As co-authors, we can address these questions in an even-handed manner. B.B. was one of the principal architects of the FBI's DNA typing programme, whereas E.S.L. was an early and vigorous critic of the lack of scientific standards, and served on the NRC committee. In a world of soundbites, we are often pegged as, respectively, a "proponent" and an "opponent" of DNA typing. Such labels greatly oversimply matters, but it is fair to say that we represent the range of scientific debate.

We recently discussed the current state of DNA typing, and could identify no remaining problem that should prevent the full use of DNA evidence in any court. What controversy existed seems to have been fully resolved by the NRC report, the TWGDAM guidelines and the extensive scientific literature. The DNA fingerprinting wars are over.

Our goal is to correct the lingering impression to the contrary. Our analysis below represents our unanimous opinions (apart from specific comments about the workings and intent of the NRC committee itself, which necessarily are based on E.S.L.'s recollection). We focus on the subject most often said to remain problematical: population genetics. Our main thesis is that the academic debate that continues to swirl about population genetic issues is rooted in a misunderstanding of the NRC report and is, in any case, of no practical consequence to the courts. We also touch on how the legal and scientific community should cope with the continuing evolution of DNA typing technology. In particular, we question whether a steady succession of *ad hoc* committees, however distinguished, is a wise solution.

Laboratory practices

The initial outcry over DNA typing standards concerned laboratory problems: poorly defined rules for declaring a match; experiments without controls; contaminated probes and samples; and sloppy interpretation of autoradiograms[1]. Although there is no evidence that these technical failings resulted in any wrongful convictions, the lack of standards seemed to be a recipe for trouble. To address these problems, the NRC committee enunciated conservative standards for each laboratory step, based on more than a decade of experience with human DNA



Peter Menzel/SPL

Comparing autoradiographs from DNA samples at Cellmark Diagnostics.

analysis. TWGDAM also developed guidelines along similar lines. Today, there is no doubt about the correct laboratory protocols to ensure reliable DNA typing results. Since the NRC report, US courts have unanimously accepted the technical reliability of DNA evidence, both in principle and in practice.

The NRC committee also highlighted the importance of laboratory accreditation, rigorous quality assurance and quality control (QA/QC) programs, and external blind proficiency tests (tests administered by persons outside the testing lab itself). The importance of these practices has been universally acknowledged, and most forensic labs follow TWGDAM's voluntary quality-control guidelines.

Population genetics

The controversy over population genetics began as a secondary issue. If DNA analysis reveals that two samples match at the loci tested, the final step is to estimate the frequency of the shared genotype in the general population, which indicates the probability that a randomly chosen person would carry this genotype. Such estimates depend on surveys of the appropriate population.

In some early cases, the rarity of genotype frequencies was greatly overstated owing to a technical error: the calculations were based on overoptimistic assumptions about the precision with which genotypes could be measured. One commercial lab, for example, reported the astronomical frequency of 1 in 738,000,000,000,000, based on a four-locus match[1]. The NRC committee easily rectified these problems by requiring consistency between the measurement precision used for forensic analysis and for population genetic estimates (a practice that the FBI, in fact, had long followed).

A subtler but more challenging issue emerged in later cases, concerning the structure of human populations. The 'product rule', used by forensic labs to calculate genotype frequencies, assumed that the individual alleles comprising a genotype could be treated as statistically independent, and their frequencies multiplied[2]. However, some population geneticists asserted that the assumption of independence was appropriate for well-mixed populations (technically, those at Hardy–Weinberg equilibrium and linkage equilibrium), but was not necessarily valid for populations with substructure. According to this argument, the frequency of a common Japanese genotype might be underestimated because the product rule ignored the fact that common Japanese alleles tend to occur together in the US Asian population. Moreover, the frequency of genotypes arising from mixed ethnic ancestry might be understated because the product rule was typically applied to separate racial databases (for the Caucasian, Black and Hispanic populations) and thus did not account for the presence of genotypes involving common alleles from different racial groups. The substructure argument became a *cause célèbre*, pitting such luminaries as Lewontin and Hartl[3] against Chakraborty and Kidd[4]. Both sides conceded that substructure could matter in principle, but many doubted that its effect could be significant in practice (see ref. 5).

The NRC committee at first attempted to settle the issue on its merits. The members agreed that the product rule was probably near the mark, but were hard pressed to say just how close. The committee considered applying formulas from theoretical population genetics based on the empirical measures of the degree of variation and admixture among and within populations. However, it concluded that there were, at the time, too few hard data about the loci used in forensic typing (most classical genetic surveys concerned protein polymorphisms, likely to be strongly influenced by natural selection) and about the precise structure of the US population. It would be too risky to base a recommendation on assumptions that might subsequently turn out to be faulty.

Thomas Caskey (Baylor College) eventually pointed the way out the quagmire when he asked, out of frustration, whether it was possible to ignore population substructure altogether. Taking up the notion, the NRC committee set out to fashion an extremely conservative rule having the virtue that it made virtually no assumptions.

The ceiling principle

The solution turned out to be quite simple. Suppose that the US population is descended from a collection of populations $P_1, P_2, \ldots, P_n$, each sufficiently old and well mixed to allow the product rule to be safely applied. Regardless of the population substructure, the multiplication rule requires only a slight modification to yield a strict upper bound on the frequency of any genotype G: for each allele in G, the allele frequency should be taken to be the maximum over the component subpopulations. In effect, the approach makes the worst-case assumption that the population may contain individuals who, for example, carry a common Caucasian allele at a locus on chromosome 2 and a common Black allele at a locus on chromosome 17. By assuming the worst, one is guaranteed to be conservative. Because it used the maximum frequency in any subpopulation, the method was dubbed the 'ceiling principle'[2].

In practice, it is unnecessary to survey every possible subpopulation. The committee concluded that the likely variation in allele frequencies could be reckoned by conducting modest surveys of 100 individuals from each of 10–15 representative subpopulations spanning the range of ethnic groups represented in the United States — such as English, Germans, Italians, Russians, Navahos, Puerto Ricans, Chinese, Japanese, Vietnamese and west Africans. Each allele frequency could then be taken to be the maximum over these subpopulations, although never less than 5%. (The latter provision was designed to deal with unexamined populations. If an allele was rare in the 10–15 subpopulations surveyed, genetic drift was not likely to have caused its frequency to drift much above 5% in other significant subpopulations.) Even in advance of detailed data about ethnic groups, the committee felt that same principle could be applied to the available racial databases (Caucasian, Black, Hispanic, Asian), although it recommended a 10% floor on allele frequencies to reflect the greater uncertainty about subpopulation variation: this slightly amended form was called the 'interim ceiling principle'. (The choices of 5% and 10% were based on the quantitative effect of genetic drift on the match odds — that is, on the reciprocal of the allele frequency — although none of this reasoning survived into the text of the final report.) The practical effect of these rules was to limit the contribution of any single locus to a factor of 50:1 odds based only on aggregate data for racial classifications and 200:1 odds based on more detailed ethnic surveys.

The ceiling principle was unabashedly conservative. It gave the benefit of every conceivable doubt to the defendant, so that it could withstand attacks from the most stubborn and creative attorneys. Some of the statistical power was sacrificed to neutralize all possible worries about population substructure.

The committee was comfortable with such a lop-sided approach, because even these extreme assumptions did not undermine the practical use of DNA fingerprinting. A four-locus match performed by forensic labs could still provide odds of 6,250,000:1. If this were not enough, two additional loci could increase the odds to more than 15,000,000,000:1.

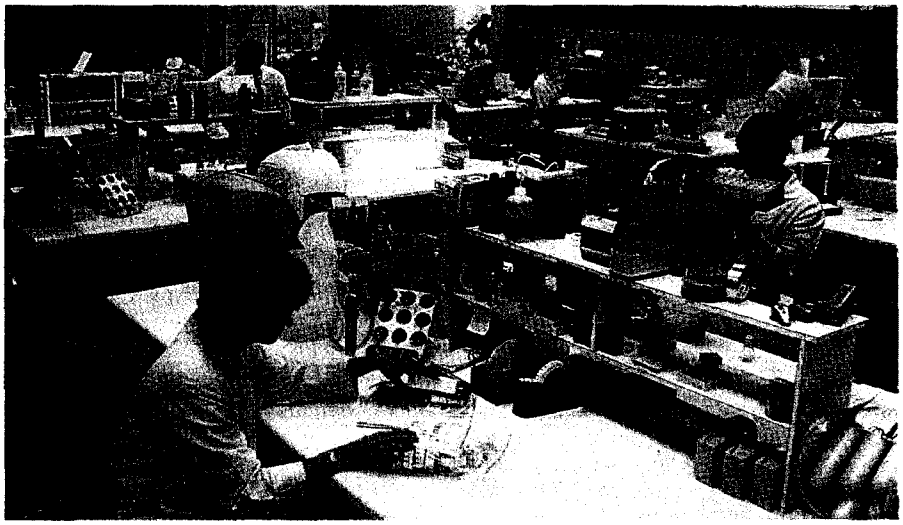
Finally, the ceiling principle was not intended to be exclusive. Expert witnesses were still free to provide their statistical "best estimate" of genotype frequencies based on the product rule. But if disagreement over such estimates arose, the ceiling principle provided an approach that all parties had to admit was biased to favour a defendant. By all rights, this seemingly solomonic solution should have ended the controversy over population genetics.

Hitting the ceiling

Surprisingly, attacks came from an unexpected quarter. Some vocal theoretical population geneticists and statisticians concluded that the committee had been too conservative. They argued that the

effect of population substructure was slight and that it would best be treated by using formulas from theoretical population genetics. The ceiling principle was accused of being clumsy and scientifically flawed. Suddenly, a new controversy over population genetics seemed to emerge[5–10].

The debate was based on a simple misunderstanding of the NRC Committee report but, with the committee disbanded, there was no easy way to address it. Moreover, the committee members had agreed to let the report speak for itself to avoid the emergence of conflicting gospels according to different members. In retrospect, this was probably an unwise decision because it has allowed a minor academic debate to snowball to the point that it threatens to undermine the use of DNA fingerprinting by suggesting that there is some problem with the use of population-genetic estimates in court.



Peter Menzel/SPL

Opting for conservative procedures — FBI serology laboratory in Washington, DC.

Six objections have been raised to the ceiling principle, which are worth briefly refuting:

(1) **The ceiling principle is premised on the flawed analysis of Lewontin and Hartl that there is significant population substructure[8].** On the contrary, the committee was quite dubious that substructure had significant effects, but felt that the possibility needed to be taken seriously rather than dismissed based on theoretical or indirect arguments. The NRC report cites the Lewontin–Hartl article[3] only twice, both times balanced against a longer list of opposing articles.

(2) **The ceiling principle is scientifically flawed because it is not used in population genetics[9,10]. Moreover, the plan to sample 10–15 representative populations is statistically unsound[8].** The choice of a statistical method necessarily depends on the dangers of overestimation versus underestimation. In forensics, there is strong agreement on the need to be conservative. By contrast, population geneticists do not need to be conservative in academic studies; they are content to err equally often on the high and low side, and thus ceiling approaches are unnecessary. However, ceiling approaches are common in a closely related genetic pursuit: the mapping of disease genes. To guard against falsely implicating or excluding a chromosomal region, human geneticists often analyse their data under worst-case conditions, such as using an unrealistically low ceiling on the penetrance of a disease gene or an unrealistically high frequency for a marker allele[11,12]. Also, some authors have complained that surveys of 100 individuals do not allow accurate estimates of the frequency of rare alleles[8]. The purpose of the suggested population studies, however, was not to estimate low frequencies, but rather to check that some alleles do not unexpectedly have extremely high frequencies (much more than 5–10%) in certain populations. For this purpose, 100 individuals is quite adequate.

(3) **The ceiling principle makes ludicrous assumptions about the possible substructure of a population[9,10].** Although the NRC report called for empirical studies of those groups that made significant contributions to the United States (such as English, Italians and Puerto Ricans), some commentators[9,10] were carried away by hyperbole — asserting that the ceiling principle assumes that "the culprit might be . . . a Lapp for one allele and a Hottentot for the other". However, if Lapp and Hottentot are replaced by Italians and Puerto Ricans, the assumption is perfectly reasonable. Indeed, it is unreasonable to assume that such genotypes don't occur in the population.

(4) **The ceiling principle is so conservative that it hampers the courtroom application of DNA fingerprinting.** Despite fears that an ultraconservative standard would clip the wings of the DNA fingerprinting, published analyses by several groups[13,14] agree that the effect is modest: Whereas the product rule typically gives four-locus genotype frequencies of about 10^{-8}–10^{-9}, the ceiling principle pares them back to about 10^{-6}–10^{-7}. That extreme assumptions have so little effect only underscores the power of DNA typing and the wisdom of taking a conservative approach.

(5) **The ceiling principle is not actually guaranteed to be conservative.** This concern appears in only a single paper[15], in which the authors point out that the conservativeness of the ceiling principle depends on the component populations $P_1, \ldots P_n$ being themselves well-mixed. The authors take great pains to construct counterexamples in the event that the populations are themselves substructured. In fact, the NRC committee considered this point to be self-evident (although failed to state it clearly enough), as can be seen from the trivial observation that the ceiling principle has no effect on genotype frequencies when one combines two identical, but substructured populations. In applying the ceiling principle in practice, the committee was confident that any residual substructure in the component subpopulations could safely be ignored in the context of such a conservative scheme.

(6) **The NRC report is causing DNA fingerprinting cases to be thrown out of court[8].** To the contrary, most courts have used the NRC report as strong evidence that, notwithstanding disagreement over the best solution, there is at least one approach that is indisputably conservative. Of the few cases in which DNA evidence has ever been rejected on population-genetic grounds, virtually all involved evidence predating the NRC report. These courts have cited the report solely for its acknowledgement that a controversy existed and was a reason for constituting the committee.

The NRC report, to be sure, has important flaws. The ceiling principle was not elegant solution, but simply a practical way to sidestep a contentious and unproductive debate. The report had more than its share of miswordings, ambiguities and errors, many of which have been corrected by a vigiliant commentator[16]. A few poorly worded sentences have been seized upon by lawyers trying to undermine the straightforward calculation of ceiling frequencies (although such arguments have not succeeded). Most important, the report failed to state clearly enough that the ceiling principle was intended as an ultra-conservative calculation, which did not bar experts from providing their own 'best estimates' based on the product rule. This failure was responsible for the major misunderstanding of the report. Ironically, it would have been easy to correct.

A law-enforcement perspective

Even as academics debated fine points, forensic scientists got on with their busi-

ness. The FBI and TWGDAM found the ceiling principle to be unnecessarily conservative, but nonetheless promptly adopted precise guidelines for implementing the ceiling principle, correctly clarifying a few ambiguous statements in the NRC report, such as which population databases to include and whether to sum adjacent bins in a frequency distribution. Forensic labs adopted a two-tiered approach, in which experts are prepared to quote both their best estimate and the conservative ceiling bound. As new population-genetic issues arise (such as how to modify or replace the ceiling principle to accommodate the less polymorphic PCR-based systems), the community is preparing to develop further guidelines. Overall, the system meets the spirit and the letter of the NRC report.

Conservative calculations have had no noticeable impact on the use of DNA evidence. In the vast majority of cases, a jury needs to know only that a particular DNA pattern is very rare to weigh it in the context of a case: the distinction between frequencies of 10^{-4}, 10^{-6} and 10^{-8} is irrelevant in the case of suspects identified by other means.

The FBI has also rapidly carried out population surveys, as recommended by the NRC committee. FBI scientists have studied more than 25 distinct subpopulations, as well as 50 separate samples from the US population[17–21]. The effort has yielded a remarkable database for examining allele frequency variation among ethnic groups. Reassuringly, the observed variation is modest for the loci used in forensic analysis and random matches are quite rare, supporting the notion that the FBI's implementation of the product rule is a reasonable best estimate. Nonetheless, the FBI has taken the scientifically sound position that it remains wiser to study new loci empirically than to assume that significant variation can never occur.

Most important, the admissibility of such DNA evidence prepared in accordance with the NRC recommendations is firmly established in virtually all US jurisdictions. In a few, the appellate courts have yet to rule formally, but there is little doubt that they will find such evidence acceptable.

Modest proposals

Although the system ain't broke, there is no shortage of proposals about how to fix it. Some academic commentators advocate a return to the product rule; others propose an approach based on the kinship statistic F_{ST}; and still others recommend an approach involving likelihood ratios that combine gel electrophoresis artefacts and population-genetic considerations into a single statistic[9,10,22–24]. Some seek to determine genotype frequencies exactly, while others prefer conservative estimates. The NRC — at the urging of the National Institute of Justice, representing the academic wing of forensic scientists — has concluded that the best solution is to constitute another *ad hoc* committee on DNA fingerprinting, composed primarily of statisticians and population geneticists.

It is easy to forget that this new debate is purely academic. The most extreme positions range over a mere two orders of magnitude: whether the population frequency of a typical four-locus genotype should be stated, for example, as 10^{-5} or 10^{-7}. The distinction is irrelevant for courtroom use.

Rehashing issues may be a harmless pastime in the academic world, but not so in a legal system that lives by the dictum *stare decisis* (let the decision stand). From the standpoint of law enforcement, it is better to have a settled, if slightly imperfect, rule than ceaselessly to quest after perfection. Already the NRC's intention to re-examine forensic DNA typing has been seized upon by some lawyers as evidence that there remain fundamental problems.

Ad hoc committees typically imagine that they will be able to accomplish their mandate with speed and finality. The original NRC study was anticipated to take one year, but required three. The idea of a second NRC panel was first floated in June 1993 with the optimistic projection that it could report by the end of that year. In fact, the committee has only just begun meeting and will probably not issue a report before late 1995. Even then, any recommendations will take 3 years to ripple through the legal system — guaranteeing that finality will not be achieved on these issues before early 1999. Despite the committee's best efforts, any new report will probably offer new opportunities for misunderstanding that will become apparent only after the panel is disbanded. And, if the new report endorses a different standard, some attorneys are sure to argue, rightly or wrongly, that differences between the reports demonstrate a lack of scientific consensus. These observations are not meant to dissuade the new NRC committee from its mission, but rather to point out the challenge facing any *ad hoc* group.

A sounder approach

The real solution is to recognize that forensic DNA typing has become a mature field and requires a more systematic approach. The NRC report anticipated that rapid evolution of technology would pose a steady succession of questions requiring attention. Its most important recommendation was the establishment of a permanent national committee on forensic DNA typing (NCFDT) to address issues as they arose. If such a committee had been appointed in 1992, it could have made short work of the population-genetics issues, by clarifying, changing or discarding the original NRC recommendations.

It is encouraging that this NRC recommendation has recently been adopted. The newly enacted DNA Identification Act of 1994 mandates the FBI to establish a DNA advisory board to recommend standards for laboratory procedures, quality assurance and proficiency testing. The act requires open meetings and broad representation, including molecular and population geneticists not affiliated with forensic laboratories, with board nominations to come from professional organizations including the National Academy of Sciences. Ideally, the panel will provide a forum for weighing important issues, including new laboratory techniques, population genetics and proficiency testing.

Most of all, the public needs to understand that the DNA fingerprinting controversy has been resolved. There is no scientific reason to doubt the accuracy of forensic DNA typing results, provided that the testing laboratory and the specific tests are on a par with currently practiced standards in the field. The scientific debates served a salutary purpose: standards were professionalized and research stimulated. But now it is time to move on. □

Eric S. Lander is in the Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA and the Department of Biology, MIT, Cambridge, Massachusetts 02139, USA. Bruce Budowle is in the Forensic Science Research and Training Center, FBI Laboratory, FBI Academy, Quantico, Virginia 22135, USA.

1. Lander, E. S. *Nature* **339**, 501–505 (1989).
2. National Research Council *DNA Technology in Forensic Science* (National Academy Press, 1992).
3. Lewontin, R. C. & Hartl, D. L. *Science* **254**, 1745–1750 (1991).
4. Chakraborty, R. & Kidd, K. K. *Science* **254**, 1735–1739 (1991).
5. *Science* **266**, 201–203 (1994).
6. Weir, B. S. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11654–11659 (1992).
7. Aldhous, P. *Science* **259**, 755–756 (1993).
8. Devlin, B., Risch, N. & Roeder, K. *Science* **259**, 748–749 (1993).
9. Morton, N. E., Collins, A. & Balasz, I. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1892–1896 (1993).
10. Morton, N. E. *Genetica* (in the press).
11. Schellenberger, G. D. *et al. Science* **258**, 668 (1992).
12. Terwilliger, J. D. & Ott, J. *Handbook of Human Genetic Linkage* (Johns Hopkins Univ. Press, 1994).
13. Duncan, G. T., Noppinger, K. & Tracy, M. *Genetica* **88**, 51–57 (1993).
14. Krane, D. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 10583–10587 (1992).
15. Slimowitz, J. R. & Cohen, J. E. *Am. J. hum. Genet.* **53**, 314–323 (1993).
16. Weir, B. S. *Am. J. hum. Genet.* **52**, 437–440 (1993).
17. Federal Bureau of Investigation *VNTR Population Data: A Worldwide Study* (1993).
18. Budowle, B., Monson, K. L., Giusti, A. M. & Brown, B. L. *J. Forensic Sci.* **39**, 319–352 (1994).
19. Budowle, B., Monson, K. L., Giusti, A. M. & Brown, B. L. *J. Forensic Sci.* **39**, 988–1008 (1994).
20. Budowle, B. *et al. J. Forensic Sci.* (in the press).
21. Huang, N. E. & Budowle, B. *J. Forensic Sci.* (in the press).
22. Morton, N. E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2556–2560 (1992).
23. Evett, I. E., Scranage, J. & Pinchin, R. *Am. J. hum. Genet.* **52**, 498–505 (1993).
24. Balding, D. J. & Nichols, R. A. *Forensic Sci. Int.* **64**, 125–140 (1994).

single man could undertake successfully this immense labour, and the resulting transcripts (photographs) would surpass the work of the most skilful painter."

As early as 1912 it was reported: "It has been demonstrated that by perpendicular exposures (from the air) the limits of buried artificial constructions, which in the ordinary procedure of reconnoitring often remain hidden, become clearly visible."

Professor Theodor Wiegand of Berlin (b. 1864) employed German aviators, during intervals in the first World War, in Syria, Palestine and Rumania, in archæological research along these lines. Almost invisibly small differences of height on the surface of the ground, indicating former fortifications or settlements, became visible as plainly as on a map, in aerial photographs lighted obliquely; remains of walls buried underground, etc., appear through variations in the character of the vegetation which covers them, depending on the season, species and conditions of growth. Wiegand published such pictures in 1920.

This technique of aerial archæological photography has been successfully elaborated since 1928, principally by O. G. S. Crawford, an English observation flier during the first World War. In the sand of the African Desert remains of buried towns have been discovered by this method.

Stereoscopic aerial photographs afforded valuable new information in the German search for the lost Königspfalz Werla (1937).

ANTHROPOLOGICAL AND ETHNOLOGICAL PHOTOGRAPHY

At the instigation of Arago and P. M. R. Serres (b. 1793), Thiesson made daguerreotypes of the Botocudes, a Brazilian Indian tribe, in 1844; in the year following Serres was able to present and discuss such photographs of comparative anthropology before the French Academy of Science, and advocated the foundation of a photographic ethnological collection.

Professor Gustav Fritsch (1838–1927) had visited South African tribes in 1860 and by the publication of his physiog-

nomical and ethnographical photographs "had opened a new era in this branch of science".

The Russians in 1862 took photographs of the different inhabitants of the Steppes of the Urals for the purpose of ethnographical study. One view gave a profile, the other a picture of the full face, also the people were shaven so that the true form of the skull should be visible.

Dr. Oidtmann, lecturer on anthropology at the Technische Hochschule, Aix-la-Chapelle, recommended a collection of pictorial statistical material for a history of the descent and development of mankind for later centuries: "systematic physiognomical photographs as anthropological documents" (1872). "Breeders of horses have long known the hereditary morphology of the horse and developed them by artificial selection. On the other hand, the knowledge of hereditary forms of human beings lags far behind . . . photography should be called upon to perpetuate the variable hereditary physiognomies of the generations." Oidtmann gave details for suitable exposures both front and side, and these were later generally adopted.

The introduction of biographical albums for every individual was proposed by the English anthropologist, Francis Galton, in 1903. He proposed that a picture should be taken every year, "a very sober picture, because the æsthetic fantasies of the camera have no scientific value at all, and the most perfect likenesses were those taken in prison."

The anthropological and ethnological section of the Paris Exposition of 1878 included a collection of photographs, chiefly coloured pictures of national costumes.

In the same year the English anthropologist, Francis Galton, created his composite portraits. Scientific expeditions from then on noted more carefully racial characteristics by photographs.

Photographs of criminals were classified on an anthropometric basis as early as 1885.

IDENTIFICATION PHOTOGRAPHY

As early as 1842, F. A. W. Netto wrote: "From a police point of view, the value of photography is inestimable; instead

of the usual description, 'Smooth forehead, ordinary nose', it is only necessary to calotype the owner of the passport and to insert his likeness in the passport."

The photographer, Dodero of Marseille (1851), pasted his portrait on visiting cards and predicted in his "Photographic Fantasies" that some day a signature alone would not suffice, but a photograph also would be required.

In 1856, an American girl, matrimonially inclined, advertised her intentions in a newspaper and requested applicants to send their daguerreotypes.

In Vienna, it was suggested that all prostitutes have health certificates bearing their portraits (1864).

A precaution adopted by the Bank of England in 1885 called for the photographing, without the person's knowledge, of any person in any degree suspicious or suspect drawing money on cheques.[1]

In Denver, Colarado, around 1907, it was a rule that drunken persons should be photographed at the Police Office and later when they had sobered off they were handed the photograph as a deterrent against future offence.

Privy Judicial Councillor Karl Theodor Odebrecht (1802–1866), in Berlin, laid the foundation, in 1864, for the use of photography in jurisprudence. He suggested that passports carry the photograph of the bearer in order to obviate exchange. This regulation was carried out for the first time on the season passes of the Berlin Photographic Exposition, in 1865, on the suggestion of H. W. Vogel. The Paris Exposition of 1867 used the same precaution.

The identification photograph was introduced only slowly in official and private documents; the Paris photographer A. Liébert (1827–1914) undertook, in 1885, to photograph all 30,000 subscribers to the newspaper, *Le Figaro*, who would present themselves for a sitting. He would deliver their portraits to them in postage-stamp size in a small card case, which carried the official stamp of the newspaper and the signature of the individual and would be recognized by officials for identification.

In 1892 all Chinese living in the United States, some 140,000 individuals, were called upon to be photographed by

[1] The Bank of England cannot trace the existence of any measure of this type.

an order in law. In the first World War similar identity cards were instituted in Europe.

Since about 1891, in Prussia, the corpse of every unidentified person has been photographed. It was proposed in France (1892) that all soldiers fallen on the battlefield, who could not be identified, be photographed; stretcher bearers were to be equipped with photographic apparatus, and properly instructed in its use.

In the Russo-Japanese War, 1904–5, every Japanese soldier carried an identity card with his photograph. The card contained also an address of next-of-kin and a prepared obituary notice so that his comrades could by this simple means communicate the news of his death.

Photography and Crime

Quite seriously, newspapers and periodicals published, in November 1839, startling information of the assistance given by "the daguerreotype as witness in a divorce case"; it was said that a husband had succeeded in photographing his spouse during a tryst without being discovered.

In June 1841 it was reported: "The worthy police will also take advantage of this method and, instead of keeping inaccurate descriptions of suspected persons, will immortalize the physiognomies of all thieves and crooks, by their own official daguerreotypist, thus establishing for ever an interesting gallery of notorious contemporaries."

In November 1841, the *Münchener Morgenblatt* reports: "The Paris police now has daguerreotypes of the features of all criminals passing through its hands, and attaches these portraits to the respective reports. When set free and suspected of a new crime, the portrait is shown to all police officers, who soon seek out their man. Daguerre certainly never dreamed that his art would be used for such a purpose."

Records of the early years show that, for instance, in 1854, the examining magistrate in Lausanne had a thief photographed, whose identity could not be established; the picture was sent to several police departments for identification. Similar cases were solved in Vienna in 1858, and at Soldin in 1860 through the assistance of a photographer.

In 1859 followed a suggestion for photographing the scene of crimes.

The London police complained in 1863 that thieves strove with all their might against being photographed.

An important book on "The Use of Photography in the Penal Process" was written by Odebrecht (1864) who proposed the following fields of application: photography of living persons (criminals), corpses discovered, scenes of crimes, documents, articles found, objects of dispute and portraits on identification cards and passports.

Rogues' galleries were first installed, about 1865, at the police headquarters of national capitals; there was one in Danzig in 1864, in Moscow it was installed in 1867; in England the photographing of criminals had been introduced generally by 1870.

Dr. Oidtmann recommended (1872) a rule "once full face, and once in profile" for portraits which were intended for research into genealogical and racial history, and the practice of the criminal police was similar.

Paris equipped its own police studio in 1874, while in Berlin, as late as 1890, photographs were made by professional photographers.

In Berlin (1877) the rogues' gallery was divided in ten groups according to the class of crime, and required the services of three officials. The album contained at the end of 1880—2,135, end of 1883—3,459 and end of 1886—3,825 records of persons photographed; in 1933, the "photographic card index" was divided into twenty-nine groups, recording 46,708 criminals. In 1932, photographs were made of 9,994 persons, 336 scenes of crime, 2,657 photographs of details of criminal clues, forty-five of corpses and many others of various subjects. The total number of prints made was 133,464.

The Prefect of the Paris Police received, between 1877 and 1882, about 50,000 portraits of criminals. In 1886 they were classified according to anthropometrical rules. It was suggested, in 1883, to photograph the ear of the criminal also, since it, unlike facial features, does not suffer any considerable change in the course of years. In 1889, Paris had about 100,000 photographs of criminals.

In Copenhagen in 1896 a complete reversal of the

normal procedure was discovered. Anarchists had got together a great collection of Danish Police officers' photographs with notes on their grade of "dangerousness."

Alphonse Bertillon (1853–1914), Paris, worked from 1882 on the identification method named after him. His pamphlet, "La Photographic Judiciaire," was published in 1890. In this work he lays down, among other instructions, exact rules for the negative (three positions of the head); his system has been universally adopted.

A photograph of the scene of a crime, as early as 1867, was held sufficient to invalidate the testimony of eye-witnesses; ten years later, the value of such photographs was recognized everywhere. In addition to the general views of a scene, the photographic record of all traces left by the crime was considered as of practical value; marks of bullets, tools, tracks, feet, blood, tyres and similar details in the picture assisted the solution of crime.

In 1887, banking establishments were equipped with flashlight apparatus, which electrically photographed persons attempting unlawful entry. It is reported that such a record at New York in 1893 led to the identification of thieves.

Since 1933, most traffic policemen in Berlin have been equipped with photographic apparatus of the size 4·5 by 6 cm, in order to record the scene of an accident before anything is changed or marks obliterated.

According to a regulation made in Prussia, about 1891, the police were instructed to photograph every unidentified corpse.

Photomicrographs of laboratory blood tests were used to ascertain whether blood was human (1877). These methods of detection were welded into a valuable range of aids to forensic investigation by the work of Dr. Paul Jeserich (1854–1927) of Berlin. Spectroscopic methods had revealed the morbid changes that could take place in blood since about 1893, and by the same means Professor Louis Lewin (1850–1929) and Erich Stenger in 1906 had detected blood in dilutions of one part in forty thousand.

Proof of forgeries or alterations of documents is a special field of legal photographic procedure; as early as 1869 it was attempted, to make visible the invisible.

COMPANY PROFILE

CELLMARK
DIAGNOSTICS

Cellmark Diagnostics is the recognized world leader in DNA FINGERPRINTING℠ service and the largest independent forensic DNA testing laboratory in the United States serving both prosecution and defense counsel needs. The company was formed in 1987 and has since analyzed thousands of samples from criminal and paternity cases from all over the United States, Canada, Puerto Rico, Guam, and South American countries. Cellmark scientists have testified in over 400 criminal cases. Cellmark Diagnostics offers complete and confidential DNA laboratory services to identify individuals using their genetic structure.

Since its inception, Cellmark Diagnostics has garnered many firsts:

* It was the first independent DNA laboratory to be accredited by the American Society of Crime Laboratory Directors - Crime Laboratory Accreditation Board (ASCLD-LAB).

* It was the first forensic DNA laboratory to use Short Tandem Repeat (STR) DNA analysis to identify military casualties for the United States Armed Forces in Operation Desert Storm. Cellmark collaborated with the Baylor College of Medicine in Houston, Texas on the project.

* It was awarded a contract by the United States Department of Defense to conduct research specifically focused on the recovery of DNA from problematic surfaces of physical evidence likely to be encountered in terrorist-related cases.

* It is the only forensic DNA laboratory to offer a national "pro bono" DNA identification service to indigent criminal defendants. Cellmark utilizes DNA technology to analyze relevant physical evidence in up to six pro bono cases per year.

Cellmark Diagnostics, Inc. is a subsidiary of Lifecodes Corporation. Located in Stamford, CT. Lifecodes is the leading manufacturer and worldwide supplier of DNA products for human identity and transplantation. Lifecodes and Cellmark Diagnostics are privately held and employee owned.

Cellmark Diagnostics, Inc.
20271 Goldenrod Lane
Suite 120
Germantown
Maryland 20876 USA

800-USA-LABS
301-428-4980
FAX: 301-428-4877

Cellmark Diagnostics, Inc. is a subsidiary of Lifecodes Corporation

FORENSICS

FEE SCHEDULE

CELLMARK DIAGNOSTICS

ASCLD/LAB ACCREDITED LABORATORY

DNA TESTING

Per Sample

RFLP TESTING:

HinfI 5 Probe Analysis	$595
HaeIII 5 Probe Analysis	$595
DNA Extraction Only (Insufficient DNA)	$275
Re-analysis of Already-Extracted Samples	$345

PCR-BASED TESTING:

DQA1 & PM	$595
STRs* (CSF1PO, TPOX & THO1 + Amelogenin)	$595
D1S80*	$595
PCR amplification only	$500

COMBINATION TESTING:

Any 2 tests	$1045
Any 3 tests	$1445
Any 4 tests	$1745

RUSH CASE FEE ..... *Additional Per Sample* $350

OTHER DNA TESTING SERVICES:

Processing of already extracted sample	
for RFLP *or* DQA1 & PM *or* STRs *or* D1S80	$350
for sexual assault mixed stains	$500
Handling fee	$100

OTHER FORENSIC SERVICES

Consulting on work done by another laboratory	$200/hour
Expert witness fees	$1,200/day + expenses for a Ph.D. $1,100/day + expenses for a non-Ph.D.
Depositions	$200/hour + expenses for a Ph.D. $150/hour + expenses for a non-Ph.D.

* This test is performed pursuant to licensing arrangements with Roche Molecular Systems, Inc. and the Perkin-Elmer Corporation.

TERMS OF PAYMENT

An invoice for the fees due will be forwarded upon completion of the case.

Terms are net 30 days. Prices are effective as of 11/1/96 and are subject to change.

Cellmark Diagnostics, Inc.
20271 Goldenrod Lane
Suite 120
Germantown
Maryland 20876 USA

800-USA-LABS
301-428-4980
FAX: 301-428-4877

Cellmark Diagnostics, Inc. is a subsidiary of Lifecodes Corporation

CELLMARK DIAGNOSTICS:
FULL SERVICE FORENSIC DNA TESTING

CELLMARK
DIAGNOSTICS

RFLP: *(5 probes) well established DNA Profiling method, statistically powerful.*

HinfI: initial screening with cocktail of 4 DNA probes for quick preliminary results.

HaeIII (FBI Protocol) test results available on disk (CODIS Lite!) for direct comparison to CODIS offender data bases.

PCR*: *most sensitive DNA test for samples too small or degraded for RFLP testing.*

PM & DQA1
STRs (CTT) & **Gender Identification** (Amelogenin)
D1S80

Only Private **ASCLD/LAB** Accredited Laboratory

Four Court-qualified Ph.D.s, testimony in over **400** cases

NO BACKLOG

BROAD CASE **ACCEPTANCE** POLICY: including evidence from sexual assault, serial and multiple suspect cases and other violent crimes, no-suspect cases, fetal tissue, urine identification, and other cases not accepted by government laboratories.

Offers **Proficiency** Testing Program Accredited by ASCLD/LAB

FREE Pre-trial Consultation

* Cellmark Diagnostics is licensed by The Perkin Elmer Corporation, under Patents owned by Hoffman-La Roche Inc., to use the PCR process to perform Forensic Analysis and Human Identification Services tests.

Cellmark Diagnostics, Inc.
20271 Goldenrod Lane
Suite 120
Germantown
Maryland 20876 USA

800-USA-LABS
301-428-4980
FAX: 301-428-4877

Cellmark Diagnostics, Inc. is a subsidiary of Lifecodes Corporation

CELLMARK
DIAGNOSTICS

20271 Goldenrod Lane · Germantown, Maryland 20876
Telephone: (301) 428-4980 (800) USA-LABS
Administration Fax: (301) 428-4877
Laboratory Fax: (301) 428-7946

FORENSIC CASE SUBMISSION

When submitting forensic cases to Cellmark Diagnostics, the cover letter should include:

1. What DNA testing (RFLP or PCR) is required and which samples are to be compared.

2. Names and phone numbers of all persons with whom Cellmark is authorized to discuss case.

3. Names and addresses of all persons to whom Cellmark should send copies of our laboratory reports.

4. Inventory of evidence being sent to Cellmark, listing exhibit number and full description of each exhibit.

5. Condition of samples, including storage history of each sample (date of collection, temperature of storage, etc.)

6. Prepayment required (or name and address to whom invoice must be sent). Please send a copy of the cover letter attached to the outer container when evidence is shipped to Cellmark Diagnostics.

7. Court dates.

8. Please put the letter or a copy of the letter on the outside of the package containing the samples.

CELLMARK
DIAGNOSTICS

20271 Goldenrod Lane · Germantown, Maryland 20876
Telephone: (301) 428-4980 (800) USA-LABS
Facsimile: (301) 428-4877

COLLECTION, PACKAGING, STORAGE, AND SHIPMENT OF EVIDENCE FOR DNA TESTS AT CELLMARK DIAGNOSTICS

COLLECTION AND PACKAGING

1. Use clean latex gloves for collecting each item of evidence. It is recommended to change gloves between the handling of different items of evidence.

2. Each item of evidence must be packaged separately.

3. Bloodstains, semen stains, and other types of stains must be thoroughly air-dried and packaged in sealed paper envelopes or paper bags. For proper chain of custody, all packages must be marked with case no., item no., and date, and must be initialed across the seals.

4. If stains must be transferred from an unmovable surface (such as a window or sidewalk), sterile cotton swabs and distilled water may be used.
   - A. Photograph the surface with a ruler before swabbing. Moisten the swab with water and shake the swab to remove the excess water.
   - B. Rub the stained area with the moist swab until all of the stain is transferred to the swab. If one swab is insufficient to collect all the stain, use additional moist swabs to collect all of the stain.
   - C. Two control swabs may also be collected as controls for other serological tests: (Swab 1) Swab an unstained area adjacent to the stained area using a moist swab. (Swab 2) Provide a moist swab with nothing else on it but the water used in the collection process.
   - D. Prepare properly marked envelopes (such as coin envelopes) or paper containers for the swabs.
   - E. Air dry the swabs without permitting the swabs to touch one another. If time requires, the swabs may be placed still moist in the envelopes until they can be transported to a place where they can be properly air dried. (This is why paper containers are preferred and glass or plastic containers should be avoided. Paper containers allow moisture to escape to prevent bacterial degradation of the DNA.)
   - F. Place swabs in appropriate separate paper containers, properly marked for identification.
   - G. Scraping dried stains should only be used instead of swabbing if the surface is perfectly smooth and the scraping will result in almost no loss of material. For example, a stain on a smooth vertical surface can be collected (after photographing with a ruler in the picture) by folding a clean sheet of paper in half and taping the top edge of the paper to the surface directly beneath the stain. With a sterile scalpel blade or unused single-edged razor blade, the stain can be scraped into the fold in the paper. Then carefully remove the paper from the surface, remove the tape, fold the paper into a packet, seal with evidence tape and initial properly.

5. Evidence which is incapable of drying such as pieces of tissue, organ, bone, liquid urine, or other biological material, should be packaged separately in an air tight container, sealed and marked properly for identification, and immediately frozen and kept stored frozen until

3. DRIED STAINS: Dried stains should be packaged separately and shipped in a sealed box or sealed envelope at room temperature by priority overnight courier service, but never over a weekend or holiday. It is preferred to ship only the stained portion of clothing rather than the entire garment. The suspected stain should be clearly marked.

4. UNDRIED TISSUE: To avoid putrifaction during shipping, tissue, organ, bone, and liquid urine samples (in plastic tubes) should be shipped in a sealed plastic container placed on abundant dry ice in styrofoam containers by priority overnight courier service, but never over a weekend or holiday.

5. LIQUID BLOOD STANDARDS: All liquid known whole blood standards should be packaged separately in styrofoam tube packages or wrapped individually in bubble wrap and secured with tape. The individually wrapped tubes of blood should be placed in a cardboard box surrounded by styrofoam chips for protection from breakage. Liquid known whole blood samples should be shipped at room temperature by priority overnight courier service, but never over a weekend or holiday. Never ship liquid blood on dry ice because the glass tubes will break.

6. BLOOD STANDARDS WITH HIV OR HEPATITIS: Blood standards from persons diagnosed with HIV or hepatitis should be shipped using a Class 6.2/95 CAN/8-2SAF-T-PAK container with the proper Federal Express shipping tag for dangerous goods. Blood standards should be shipped at room temperature by priority overnight courier service, but never over a weekend or holiday. Never ship liquid blood on dry ice because the glass tubes will break.

   Some packaging suppliers for the 6.2 substances including blood infected with HIV or hepatitis are listed below.

   - Source Packaging (sells SAF-T-PAK and offers 6.2 specialty training. Contact Rick Ferris (401) 738-7733.)
   - Federal Industries (sells SAF-T-PAK, (800) 523-9033)
   - Labelmaster (3M pack and others pending, (800) 621-5800)
   - All Pak (attention: Bill Barger, (800) 245-2283)

IF YOU HAVE ANY QUESTIONS, PLEASE CALL 1-800-USA-LABS

FAX TRANSMISSION

CELLMARK DIAGNOSTICS, INC.

20271 GOLDENROD LANE
GERMANTOWN, MARYLAND 20876-4064
TELEPHONE: (301) 428-4960
FAX: (301) 428-4677

To: Mr. Simpson Garfinkle **Date:** July 21, 1997

Fax #: 508-696-8989 **Pages:** 5, including this cover sheet.

From: Mark Stolorow

Subject: Sample Laboratory Reports

COMMENTS:

Please find attached samples of laboratory reports such as those we would issue at Cellmark Diagnostics. I hope these are helpful.

Please do not duplicate these documents or publish them. They are intended for instructing clients and are fairly sterile but we try to minimize the chances of our report formats from falling into the hands of those who might use them improperly. Your cooperation is appreciated.

Please call if you have additional questions.

Thank you.

-Mark

SAMPLE REPORT
Example of Match

REPORT OF LABORATORY EXAMINATION
OCTOBER 1, 1989

Detective Robert Smith
Johnstown Police Department
Johnstown, Maryland 20202

Re: Your Case No. 89-1234
Cellmark Case No. F890000

EXHIBITS:

The exhibits listed below were received for analysis on September 1, 1989:

Item #1 Sealed envelope labelled "vaginal swabs, victim" containing two (2) swabs

Item #2 Sealed envelope labelled "blood, victim" containing one (1) lavender top tube of blood labelled "blood, victim"

Item #3 Sealed envelope labelled "blood, suspect" containing one (1) lavender top tube of blood labelled "blood, suspect"

RESULTS:

DNA was extracted and DNA banding patterns were obtained from the items listed above using the restriction enzyme Hinf I and four single locus probes (MS1, MS31, MS43, g3). The DNA banding pattern obtained from the vaginal swabs (Item #1) contained the DNA banding pattern obtained from the blood sample labelled "blood, victim" (Item #2) and a second DNA banding pattern. This second DNA banding pattern matches the DNA banding pattern obtained from the blood sample labelled "blood, suspect" (Item #3).

CONCLUSION:

The frequency of the DNA banding pattern obtained from the vaginal swabs and the blood sample labelled "blood, suspect" will be provided in a subsequent report.

Staff Molecular Biologist Ph.D.

The DNA banding pattern obtained from the vaginal swabs received on [redacted] contains a DNA banding pattern which matches the DNA banding pattern obtained from the blood swatch labelled [redacted] and a second DNA banding pattern. This second DNA banding pattern matches the DNA banding pattern obtained from the blood swatch labelled [redacted]

CONCLUSIONS:

[redacted]

Using the five single-locus probes sequentially, the approximate frequencies in the Caucasian, African American, and Western Hispanic populations of [redacted] the DNA banding pattern obtained from the vaginal swabs and the blood swatch labelled [redacted] are as follows:

Population Database	Frequency
Caucasian	1 in 3.1 trillion
African American	1 in 170 billion
Western Hispanic	1 in 830 billion

Charlotte J. Word, Ph.D.
Deputy Laboratory Director

Paula J. Yates
Forensics Supervisor

SAMPLE REPORT
Example of No Conclusion

REPORT OF LABORATORY EXAMINATION
OCTOBER 1, 1989

Detective Robert Smith
Johnstown Police Department
Johnstown, Maryland 20202

Re: Your Case No. 89-1234
Cellmark Case No. F890000

EXHIBITS:

The exhibit listed below was received for analysis on September 1, 1989:

Item #1 Sealed envelope labelled "vaginal swabs, victim" containing two (2) swabs

RESULTS:

An insufficient quantity of high molecular weight DNA was obtained from the vaginal swabs (Item #1) to continue testing.

CONCLUSION:

No conclusion can be made concerning the vaginal swabs.

Staff Molecular Biologist Ph.D.