# Genetic fingerprinting

Alec J Jeffreys

**Southern**, *n.*[2] Used attrib. (chiefly in *Southern blot*, *blotting*: see *BLOT *n.*[1] 1 *f*, *BLOTTING *vbl. n.* 4) with reference to a technique for the identification of specific nucleotide sequences in DNA, in which fragments separated on a gel are transferred directly to a second medium on which assay by hybridization, etc., may be carried out.

**Genetic fingerprinting**, the obtaining or comparing of genetic fingerprints for identification; *spec.* the comparison of DNA in a person's blood with that identified in matter found at the scene of a crime, etc.

– Oxford English Dictionary

There can be few higher accolades in science than making it into the Oxford English Dictionary—except perhaps receiving the Lasker Award! So how did my involvement in DNA and forensics come about? The following is a brief personal account of an exploration into human genetic variability at the most fundamental level of all, namely variation in DNA, and of a wholly unanticipated excursion into the extraordinary world of forensic medicine and the law.

I think I was born a scientist, perhaps thanks to genes inherited from my father and grandfather, both prolific inventors with the latter developing the Jeffreys Three-dimensional Photosculpture Process that was all the rage in London in the 1930s. I was active in science by the age of eight, thanks to my father's gift of a microscope (**Fig. 1**) and a remarkably dangerous chemistry set (I still bear the scars). These triggered an abiding interest in biology and chemistry and led directly to me pursuing a degree in biochemistry at the University of Oxford, starting in 1968. But genetics soon emerged as my passion—even in those distant days, it was clear that penetrating the human genome was going to be one of the greatest challenges facing biology. So after my doctoral studies on human genetics at Oxford, I moved to the University of Amsterdam in 1975 and, by good fortune, teamed up with Richard Flavell on a project to purify a single-copy mammalian gene—the gene encoding rabbit β-globin—by mRNA hybridization. The problem was how to monitor purification. The answer was provided by Ed Southern and his blots (incidentally, and with typical modesty, Ed never calls them Southerns but generally DNA transfers), which we showed, much to our surprise, were capable of detecting single-copy genes in complex genomes. This led to the first physical map of a mammalian gene[1] and one of the first descriptions of introns[2].

## Human DNA variation

In 1977 I moved to the Department of Genetics at the University of Leicester and, at the tender age of 27, was faced with the thorny issue of where to go next with this arcane new science of genomics. The choices were endless but the answer, to me, was fairly obvious—why not try to marry together genomics with the classic discipline of human genetics and try to detect heritable variation directly in human DNA? Within months we had found our first variant in the form of a restriction fragment–length polymorphism (RFLP) and had soon surveyed the β-globin gene cluster (at the time one of the few accessible regions of the human genome) for RFLP abundance[3]. In a wild extrapolation from the cluster to the entire genome, I predicted the existence of 15 million single nucleotide polymorphisms (SNPs; actually I predicted 30 million but I made a twofold error—no one's perfect!). This prediction, based on the tiniest of surveys, is remarkably close to the current inventory of 10 million human SNPs logged into dbSNP. Beginner's luck?

## Hypervariable DNA

These RFLPs provided human geneticists with a vast new range of mendelian markers and led directly to the emergence of the first human linkage maps and to the detection of inherited disease loci by linkage. But RFLP assays for SNPs were cumbersome and generally yielded relatively uninformative biallelic markers. Spurred on by a report in 1980 from Arlene Wyman and Ray White of the accidental discovery of a hypervariable multiallelic locus of unknown physical basis[4], we turned our attention to a search for other such informative loci. Intuitively, we felt that tandem-repeat DNA loci should be a rich source of such markers. This hunch proved correct, with reports soon appearing on variable tandem repeats discovered by chance in the α-globin gene cluster[5] and near the gene encoding insulin[6]. The problem now was how to develop generic methods for isolating these loci—which we subsequently dubbed minisatellites—from complex genomes.

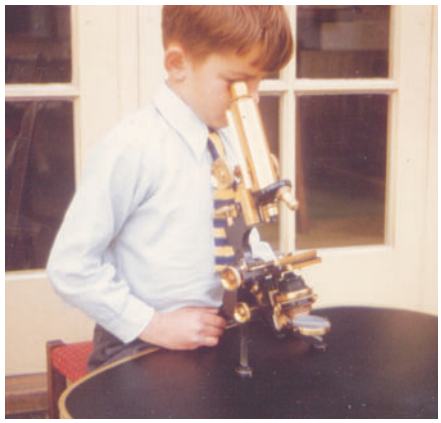

**Figure 1** Alec Jeffreys at age eight. A budding scientist not yet familiar with the focal plane.

The answer came unexpectedly from our work on the evolution of the globin gene family, including the distant relative myoglobin. Inside the gene encoding human myoglobin was a minisatellite. Curiously, there were sequence similarities between its repeat unit and the repeats of the few other minisatellites described at the time. To explore this further, we screened a human genomic library with the myoglobin minisatellite and successfully isolated additional variable loci. DNA sequence analysis revealed a short (10–15 basepairs) motif shared by these minisatellites[7] (**Fig. 2**).

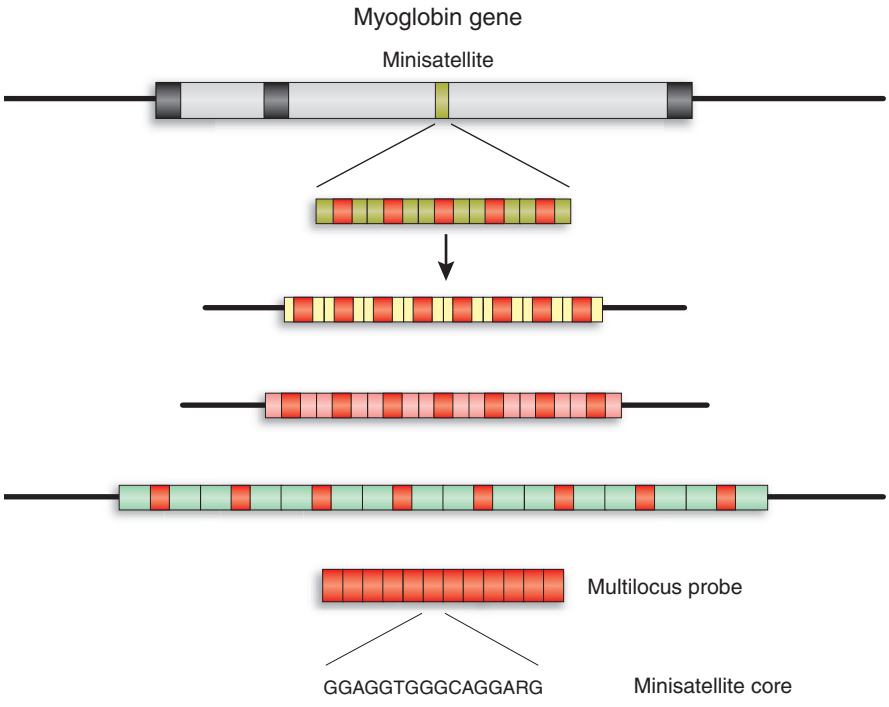


**Figure 2** The minisatellite core sequence. Cross-hybridization of the myoglobin minisatellite to other human minisatellites allowed definition of a sequence motif shared by these loci and the design of multilocus minisatellite hybridization probes[7].

Even today, the significance of this core sequence remains unknown, though presumably it predisposes these loci to DNA duplications. But its existence immediately suggested a general method for isolating hypervariable loci, namely by hybridization with probes consisting of repeats of the core sequence itself.

## DNA fingerprints

We tested this idea of detecting multiple hypervariable loci by hybridizing a core repeat probe to an arbitrarily chosen Southern blot carrying DNA from a family group plus a range of DNAs from various nonhuman species (**Fig. 3**). Although the autoradiograph was indistinct and messy, the results were amazing—emerging from the gloom were what seemed to be highly variable profiles of DNA that looked as though they were simply inherited in the family. The penny dropped almost immediately—we had accidentally stumbled upon a DNA method with potential for individual identification. Thus, on the morning of Monday, September 10, 1984, DNA fingerprinting was born. Within minutes, we had drawn up a list of possible applications, including criminal investigations, paternity disputes, zygosity testing in twins, monitoring transplants, wildlife forensics, conservation biology and the like. My wife Sue added another that evening—immigration disputes.

## DNA fingerprinting in practice

We were now faced with two major challenges. First, could we improve the technology? Second, would anyone ever take notice of this obscure new approach to identification?

We soon improved the method to the point at which rich and highly informative DNA fingerprints could be obtained from human DNA[8]. Family analyses showed that these patterns were derived from many highly variable loci dispersed throughout the human genome and that each minisatellite band could, at least as a first approximation, be treated as a statistically independent characteristic. Comparison of different people revealed huge levels of variability, even between first-degree relatives, but, as expected, identity between monozygotic twins.

The second problem was solved by the press. Our first publication on DNA fingerprinting was reported in a British national newspaper and came to the attention of Sheona York, a lawyer working at a community law center in London and representing a UK family of Ghanaian origin embroiled in an immigration case centered on the disputed identity of one of the boys. The family had been subjected to the standard genetic tests of the time, namely blood group, serological and allozyme typing. These indicated some relationship between the mother and boy in dispute but could not prove a mother-son relationship over alternatives such as aunt-nephew, and the boy thus faced the threat of deportation from the UK. At the lawyer's request, we therefore used DNA fingerprinting to test blood samples from the family in the spring of 1985 (**Fig. 4**). These tests showed that the boy was a member of the immediate family and established the power of DNA typing for positive identification rather than just for exclusion[9]. The British Home Office dropped the case against the boy and he remained with his family in London as a UK citizen. To my knowledge, this was the first case ever resolved by DNA.

Other applications followed swiftly. The first paternity case in the summer of 1985 led to DNA evidence being considered in a UK Magistrate's Court, the first time that DNA had entered a court of law. Completion of a pilot study with the Home Office on immigration cases, primarily from the Indian subcontinent, established the utility of DNA fingerprinting in resolving cases involving a UK sponsor attempting to bring his wife and children into the UK in the absence of adequate documentary evidence (and incidentally showed some interesting discrepancies between DNA evidence and the decisions of Entry Clearance Officers who had conducted family investigations, village interviews and the like). Zygosity testing appeared on UK television in a popular family show on which we established that twin sisters who were unsure as to their status were indeed monozygotic twins. The audience of over 10 million viewers probably represented at the time the largest public exposure ever to the concept of DNA technology. Nonhuman applications soon emerged, ranging from studies of mating behavior in house sparrows (like humans, they indulge in spouse-swapping[10]) to the verification of Dolly the sheep as an authentic clone[11] and the reconstruction of the entire pedigree of a colony of Waldrapp ibises at the Zurich Zoo[12], identifying instances of incestuous mating and helping zookeepers optimize outbreeding in this endangered species.

## DNA profiling

Although Peter Gill at the Home Office Forensic Science Service and I soon established that DNA could survive remarkably well in forensic specimens such as blood and semen stains[13], it was clear that DNA fingerprints were too complex and required too much high-quality DNA for forensic investigations. The solution to this problem rapidly emerged from our work on cloning individual minisatellites from DNA fingerprints to generate highly variable single-locus marker systems[14]. Some of these single-locus probes detected extraordinarily variable loci to produce much simpler two-band two-allele

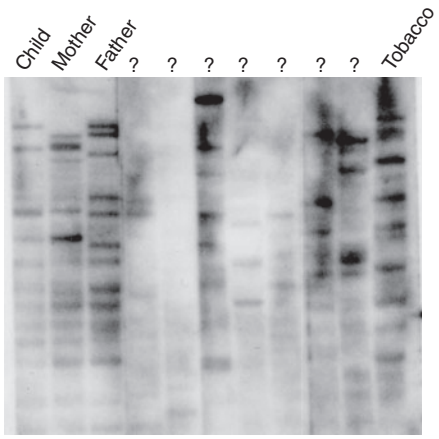


**Figure 3** The first DNA fingerprints. The autoradiograph shows a human family trio plus a range of DNAs from various species including baboon, lemur, seal, cow, mouse, rat, frog and tobacco. This autoradiograph and the accompanying lab notes were sold at auction for charity years ago (incidentally for £180, which I thought was rather good at the time) and unfortunately I forgot to keep copies and cannot recall what some of the DNAs are.

hybridization patterns on Southern blots of human DNA, and with considerably less DNA than was needed for DNA fingerprinting. It was immediately obvious that these DNA profiles had great potential for forensic investigations—the profiles could be entered into a computer database as estimated allele lengths to facilitate DNA comparisons, and mixed DNA samples could be readily identified as such by the presence of more than two alleles.

DNA profiling saw its forensic debut in 1986 with the Enderby murder case in Leicestershire, in which we showed that a prime suspect of the rape and murder of two schoolgirls did not match forensic DNA evidence[14] (**Fig. 5**). Given his confession to one of the murders plus additional circumstantial evidence, I have little doubt that without this DNA evidence he would have been convicted. Thus, the first use of DNA in a criminal investigation led to the establishment of innocence, not guilt. This aspect of DNA testing is crucial and has led to initiatives such as the Innocence Project in the US, which has championed systematic post-conviction DNA testing and has to date exonerated 159 individuals who have been wrongfully convicted and who have served years or even decades in prison, in some instances on death row. Sadly, there is, to my knowledge, no equivalent initiative in the UK.

The Enderby murder case was brought to a successful conclusion after a DNA-based manhunt in the local community and the unmasking of the true perpetrator, who otherwise would most likely have killed again[15]. After this case, DNA profiling rapidly spread worldwide to become the forensic DNA testing system of choice. However, early court cases exposed some lack of rigor in the practice and interpretation of forensic DNA[16] and led to detailed reviews by the US National Research Council into forensic DNA testing, which identified mechanisms to maximize the reliability of this extraordinarily powerful new evidence. Never has a field of forensic science been subjected to such a detailed and probing review, and it is a testimony to forensic DNA that it survived these challenges to emerge as a robust, reliable, validated technology that soon lay at the heart of countless criminal investigations.

## Amplifying the evidence

DNA profiling is powerful but has limitations; most forensic specimens yield too little intact DNA for typing. The solution was provided in the late 1980s by PCR and by the development of alternative tandem-repeat DNA markers much shorter than kilobase-long minisatellites. These microsatellites or simple tandem repeat (STR) markers[17,18] have alleles just 100 bp or so in length, which are easily amplified and ideally suited to typing trace or degraded DNA. Our first foray into STR typing was an investigation with Erika Hagelberg in 1990 into skeletal remains suspected as being those of Josef Mengele, the Auschwitz concentration camp doctor. Bone DNA profiles using $(CA)_n$ repeat loci were successfully amplified from nanogram amounts of highly degraded human DNA despite the presence of a huge excess of nonhuman (presumably bacterial) DNA. Comparison with blood DNA profiles of Mengele's wife and son (**Fig. 6**) established with good likelihood that the remains were authentic[19] and allowed a major war-crime investigation to be brought to a conclusion.

Surprisingly (to me at least), it took another five years before STR typing superseded the much more cumbersome and less sensitive DNA profiling and came to dominate the field. Developments in human genetics improved the typing systems, with a shift to tetranucleotide repeat STRs[20] that gave much less stuttered profiles than those in **Figure 6**, the introduction of multiplex PCR allowing multiple STRs to be typed simultaneously, and the use of fluorescently labeled PCR primers enabling STR profiles to be recovered, interpreted and entered into a database in much the same way as in automated DNA sequencing[21]. These developments led in April 1995 to Michael Howard, at the time the British Home Secretary, announcing the launch of the National DNA Database (NDNAD), the first integrated national database of criminal DNA profiles. NDNAD has been spectacularly successful as a crime-fighting tool and the current statistics are sobering: as of July 2005 it holds over 2,900,000 DNA profiles (plus physical DNA samples) from individuals and unsolved casework, and has to date been used to identify 630,000 suspects by matches between crime-scene DNA and NDNAD records, as well as 40,000 instances of matches between different crime-scene samples, giving the police crucial leads in, for example, instances of serial rape. And it is not just serious crime that is being investigated—even volume crime such as car theft and burglary fall within DNA's scope, given that trace DNA can be typed even from handled objects such as the steering wheel of a stolen car[22]. Half of UK crime-scene samples yielding DNA now produce suspect matches on NDNAD.

Many countries have initiated their own DNA databases, though with widely differing criteria for DNA inclusion and retention[23]. The UK database is still by far the largest both absolutely and per capita. And it is no longer restricted to convicted individuals—because of the Criminal Justice Act of 2003, suspects cleared in an investigation can now be retained indefinitely on NDNAD. The implicit assumption

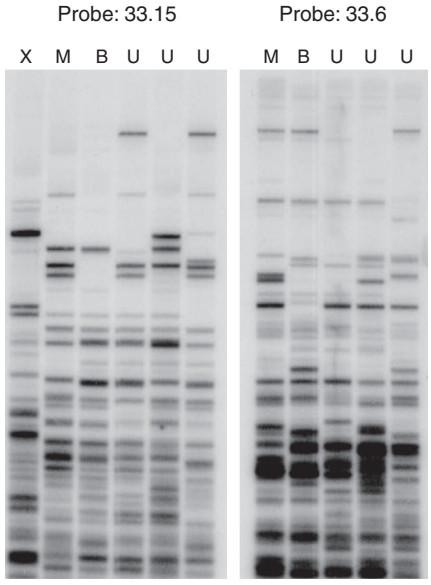


**Figure 4** The first application of DNA fingerprinting—an immigration case. Two multilocus DNA fingerprinting probes chosen to detect different sets of minisatellites were used on the mother (M), her three undisputed children (U) and the boy in dispute (B). The father was not available and X is me. All bands in B can be traced back to M or to a paternal character in one or more Us. The likelihood ratio in favor of B being a son rather than unrelated to M is $5 \times 10^8$:1, with a corresponding likelihood ratio of $3 \times 10^4$:1 for a mother-son relationship rather than aunt-nephew[9].

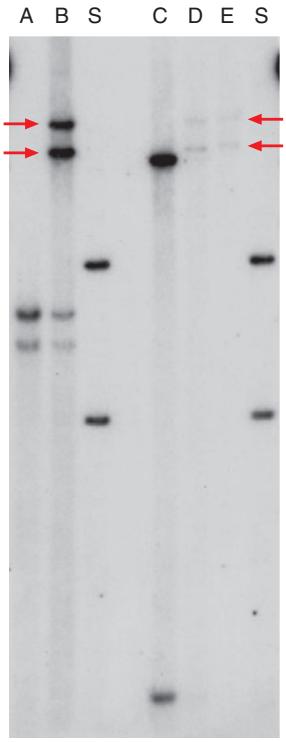


**Figure 5** The first application of DNA profiling—the Enderby murder case. A single locus minisatellite probe was used to analyze the following DNAs: A, hair roots taken post-mortem from the first victim; B, mixed semen and vaginal fluid from the first victim; C, blood taken post-mortem from the second victim, D, vaginal swab from the second victim; E, semen stain on clothing from the second victim; S, blood from the prime suspect. Semen alleles (indicated by arrows) not attributable to the victims appear shared across both murders, but do not match the suspect[14].

seems to be that these individuals, now numbering in the hundreds of thousands, are more likely than others to commit crime in the future and thus, retaining their DNA will facilitate crime detection. To my knowledge, this assumption has yet to be supported by any statistical evidence. Further, this extension of NDNAD to a subset of the general population that is likely to be skewed socioeconomically and ethnically raises issues of civil liberties and genetic discrimination. Another interesting but potentially worrisome extension to the use of NDNAD is familial searching, in which a crime-scene DNA profile that does not match anyone on the database is used to search for potential relatives of the perpetrator[23]. To date, this has been used successfully in a few cases, but does raise issues concerning the dual use of DNA both to identify repeat offenders and to inculpate relatives, as well as its potential to disclose unsuspected family relationships (or lack thereof).

## The future

Despite the power of DNA databases, many criminal cases do not yield matches with individuals whose DNA profiles are stored in the database. In such cases, there is interest in recovering physical information from DNA to help police in their investigations. Some characteristics—gender, red hair and, to some extent, eye color and ethnic origin—can be typed by DNA[23]. The intriguing notion that surnames of male perpetrators might be traced by correlations between names and Y chromosomal markers (both paternally inherited in many societies) is sadly weakened by poor levels of correlation for many names as the result of nonpaternity, adoption and multiple historical origins of some names[24]. More fanciful speculations that clues about stature and facial appearance could be gleaned from DNA variation await establishment of their genetic basis, which may well prove too complex to yield simple readouts from forensic DNA. There are also issues of genetic privacy—although the currently used batteries of forensic STR markers carry virtually no information other than identity, kinship and (weakly) ethnic origin, the same is unlikely to be true for phenotypic markers for which unexpected and unwarranted disclosure of deleterious alleles, of major significance to an individual and relatives, would represent a serious invasion of genetic privacy.

The alternative approach would be to develop DNA databases of entire populations[25]. Although there is no technical reason why such databases could not be produced, this proposal is costly and does raise huge civil liberty issues that may well make it totally unacceptable. But before dismissing the suggestion out of hand, it does merit reflection. Consider mass disasters such as the terrorist attacks of September 11, 2001, the Asian tsunami and the recent bombings in Bali, Madrid and London. DNA typing had, and continues to have, a central role in victim identification, but is hampered by the lack of reference DNA samples, which forces investigators into intrusive quests for potential victims' DNA samples (from toothbrushes, etc.) and requests for DNA samples from relatives, adding to the misery of already terrible investigations. There is therefore an argument that retaining your DNA profile as a personal certificate of identity makes sense for you, your family and society. One vision might therefore be of a comprehensive DNA database held by an agency separate from the police, with DNA profiles and personal data kept separately and with additional security afforded by encrypting DNA profiles into personal identification numbers (DNA PINs) to eliminate any residual information about kinship, ethnicity and so forth. Police would have access only to the DNA PIN database, to further criminal investigations, with strict

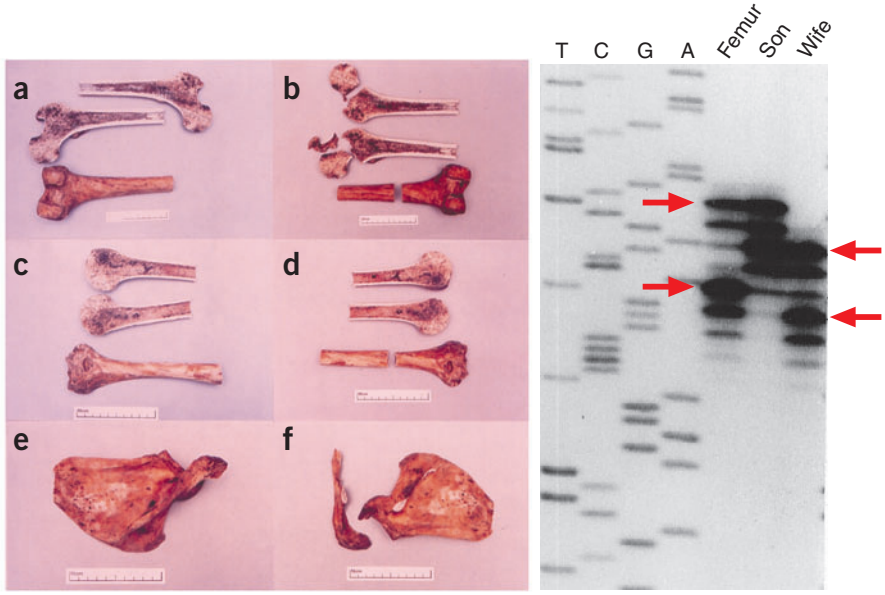


**Figure 6** The identification of Josef Mengele by bone STR typing. Some of the skeletal remains exhumed from Brazil are shown, together with PCR profiles of a $(CA)_n$ STR from femur DNA and from Mengele's son and wife. These dinucleotide STRs generate stutter bands, but the original alleles (indicated by arrows) can be readily identified. The paternal allele in the son is present in the bone DNA. Similar matches over 10 different STR loci provided a combined likelihood ratio of >1,800:1 of the remains being from Mengele rather than from some unrelated person[19].

court control over how DNA PIN matches are handled in subsequent investigations.

The future will therefore be one of steady expansion of criminal DNA databases with increasing integration internationally, and possibly extension to more comprehensive population databasing, coupled with the development of new DNA typing platforms that could allow miniaturization ('lab on a chip') and greatly accelerate testing with the potential for analyzing DNA at the crime scene. If the time for testing can be reduced to seconds—perhaps by imaging SNPs on single DNA molecules—then an entirely new dimension to DNA typing might emerge in the field of security, with DNA PINs serving as true PINs in everything from immigration clearance to credit card transactions.

## Concluding remarks

DNA typing has come a long way since its accidental beginnings 21 years ago and has had a profound impact on individuals, society and the law, having already directly touched the lives of millions of people worldwide. But there is much more to DNA fingerprinting than identification. Minisatellites have provided us with some of the most variable and unstable loci in the human genome, and have proven superb for analyzing processes of repeat DNA instability in the human germline[26]. These studies have led to the development of new single–DNA molecule approaches to investigating the dynamics of human germline mutation and recombination[27], the two great drivers of human DNA diversity, and have also provided new, highly efficient systems for analyzing environmental factors such as ionizing radiation that might have an impact on *de novo* heritable mutation in humans[28]. But these are stories for another day.



ACKNOWLEDGMENTS

I extend a huge thanks to my father and late mother for a great childhood and for enduring endless experimental misery, also to my wife and daughters for constant support and patience. There are too many friends and colleagues who have contributed over the years to name individually, but a special thanks must go to V. Wilson for her key work in the initial phases of developing DNA fingerprinting and to I. Patel for starting the first mass application of DNA testing, in immigration disputes. I am greatly indebted to the late G. Roderick of the Lister Institute of Preventive Medicine whose patience, friendship and expertise proved so valuable during the early phases of commercializing DNA testing, and to all my friends at Cellmark Diagnostics who turned a bit of academic eccentricity into something truly of value. And finally, I extend my gratitude to the Lasker Foundation that has done me such a huge honor.

1. Jeffreys, A.J. & Flavell, R.A. A physical map of the DNA regions flanking the rabbit β-globin gene. *Cell* **12**, 429–439 (1977).
2. Jeffreys, A.J. & Flavell, R.A. The rabbit β-globin gene contains a large insert in the coding sequence. *Cell* **12**, 1097–1108 (1977).
3. Jeffreys, A.J. DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes of man. *Cell* **18**, 1–10 (1979).
4. Wyman, A.R. & White, R. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77**, 6754–6758 (1980).
5. Goodbourn, S.E., Higgs, D.R., Clegg, J.B. & Weatherall, D.J. Molecular basis of length polymorphism in the human zeta-globin gene complex. *Proc. Natl. Acad. Sci. USA* **80**, 5022–5026 (1983).
6. Bell, G.I., Selby, M.J. & Rutter, W.J. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* **295**, 31–35 (1982).
7. Jeffreys, A.J., Wilson, V. & Thein, S.L. Hypervariable "minisatellite" regions in human DNA. *Nature* **314**, 67–74 (1985).
8. Jeffreys, A.J., Wilson, V. & Thein, S.L. Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76–79 (1985).
9. Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. Positive identification of an immigration test case using DNA fingerprints. *Nature* **317**, 818–819 (1985).
10. Wetton, J.H., Carter, R.E., Parkin, D.T. & Walters, D. Demographic study of a wild house sparrow population by DNA fingerprinting. *Nature* **327**, 147–149.
11. Signer, E.N. *et al.* DNA fingerprinting Dolly. *Nature* **394**, 329–330 (1998).
12. Signer, E.N., Schmidt, C.R. & Jeffreys, A.J. DNA variability and parentage testing in captive Waldrapp ibises. *Mol. Ecol.* **3**, 291–300 (1994).
13. Gill, P., Jeffreys, A.J. & Werrett, D.J. Forensic application of DNA 'fingerprints'. *Nature* **318**, 577–579 (1985).
14. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A.J. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**, 269–288 (1987).
15. Wambaugh, J. *The Blooding* (Bantam, London, 1989).
16. Lander, E.S. DNA fingerprinting on trial. *Nature* **339**, 501–505 (1989).
17. Litt, M. & Luty, J.A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397–401 (1989).
18. Weber, J.L. & May, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396 (1989).
19. Jeffreys, A.J., Allen, M., Hagelberg, E. & Sonnberg, A. Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Sci. Int.* **56**, 65–76 (1992).
20. Edwards, A., Civitello, A., Hammond, H.A. & Caskey, C.T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**, 746–756 (1991).
21. Kimpton, C.P. *et al.* Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl.* **3**, 13–22 (1993).
22. van Oorschot, R.A.H. and Jones, M.K. DNA fingerprints from fingerprints. *Nature* **387**, 767 (1997).
23. Jobling, M.A. & Gill, P. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* **5**, 739–751 (2004).
24. Jobling, M.A. In the name of the father: surnames and genetics. *Trends Genet.* **17**, 353–357 (2001).
25. Williamson, R. & Duncan, R. DNA testing for all. *Nature* **418**, 585–586 (2002).
26. Jeffreys, A.J. *et al.* Human minisatellites, repeat DNA instability and meiotic recombination. *Electrophoresis* **20**, 1665–1675 (1999).
27. Kauppi, L., Jeffreys, A.J. & Keeney, S. Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.* **5**, 413–424 (2004).
28. Dubrova, Y.E. Germline mutation induction at mouse and human tandem repeat DNA loci. *Adv. Exp. Med. Biol.* **518**, 115–129 (2003).

***Alec J. Jeffreys is in the Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. e-mail: ajj@le.ac.uk***