An overview of the use of DNA evidence in South African criminal courts

LIRIEKA MEINTJES-VAN DER WALT*

ABSTRACT
DNA evidence differs from other forensic evidence in that it developed in the scientific arena and not with the exclusive objective of being used for identification and detection of criminal perpetrators. In order to understand the forensic significance of DNA evidence in South Africa today, this article explains different aspects of the use of DNA in criminal cases. The article explains the nature of DNA evidence and genes and their relationship to individuality. The technology of DNA analysis and its application to DNA profiling is described. The power of DNA as a forensic tool is evaluated by looking at the laboratory practices within the SAPS Biology laboratories. Interpretation of the data generated is discussed as well as potential future developments of the forensic use of DNA.

Introduction
There are few, if any, techniques in the history of forensic science that have been more thoroughly scrutinized, tested and validated than forensic DNA testing. However, it is essential that lawyers are aware of the science underlying DNA evidence, the latest developments in this field, and the possible grounds upon which challenges to the admissibility and weight of DNA evidence can be made. It should, however, be made clear that the techniques used in DNA evidence are well established and scientifically validated.¹

This article deals with the application of DNA profiling to criminal investigations. Forensic DNA analysis involves the intersection of molecular biology, genetics and statistical analysis. The biological principles that underlie DNA testing are briefly explained. This is followed by an overview of current DNA testing as used by the South African Police Service Forensic Science Laboratories. Quality control and quality assurance procedures in the South African Police Service Forensic Science Laboratories (SAPS FSL) are detailed and research reports on quality assurance procedures in the laboratory are outlined. The debate concerning the interpretation of laboratory results is discussed. Problems and issues encountered in the context of DNA evidence are discussed and the significance of a match or inclusion result is then examined.

**Uses of DNA profiling in criminal investigations**

| To determine the identity of the perpetrator | Comparing a profile derived from semen in a rape victim's vagina with the suspect's profile. Male on male rapes are also investigated — although less frequently. |
| To determine whether a suspect had prior contact with the victim | Comparing a profile derived from bodily samples found on the suspect's body with a victim's profile, or vice versa. |
| To determine the identity of the victim | Comparing a profile from an unidentified person or corpse with a known person's profile. |
| To infer the common involvement of one person in separate crimes | Comparing profiles in two crime scene samples. |
| To confirm or negate a suspicion | Comparing the profile of a suspect with a profile derived from a sample found on or in the victim's body, or vice versa. |

By proving identification DNA can corroborate other elements of substantive crimes. Questions such as who, what, when, where and to a lesser degree why, can be answered by drawing inferences from the source, location or type of DNA evidence found during an investigation.

The potential meaning derived from the source, location and type of DNA evidence

| Where was DNA found? | If a DNA sample matching the victim is found in an accused's home, this evidence can negate the accused's version that the victim was never there. If DNA, such as a DNA sample extracted from blood, is recovered from a specific place then such evidence may corroborate the victim’s description of having been attacked in that place. |
| How can DNA evidence help determine what happened during a crime? | Saliva samples found under a bed may indicate that the accused was hiding there before the victim was attacked. Fingernail scrapings of the victim’s skin found under the victim’s fingernails, combined with scratches on his or her neck, may corroborate evidence of the victim’s attempt to remove a ligature from around his or her throat. |
How the location of DNA samples can assist in explaining the sequence of events or when a specific incident occurred.

How DNA can provide circumstantial evidence which could suggest the intent to commit the crime.

Impeaching the accused's version of events.

Supporting the complainant's testimony.

Finding a victim's DNA in a blood sample taken from the accused's weapon could help explain why the victim acceded to the accused's demands.

DNA evidence taken from the inside of a balaclava could contribute to the argument that the accused had intent. DNA evidence found inside a victim's house when the accused is before the court on housebreaking charges, may be consistent with the absence of consent by the home owner.

Often in sexual offences, the accused will deny knowing the victim. Confronted with his or her DNA evidence, the accused's defence may change, for example, to consent.

Where DNA from the accused is recovered from an article of clothing described by a witness, such evidence may contribute to the accuracy of the witness's testimony.

The use of DNA evidence at trials in other jurisdictions has often resulted in lawyers raising case-specific questions about issues such as the laboratory procedures used, the techniques for declaring a DNA match and problems associated with statistical calculations. This article will examine these arguments and, in conjunction with experts from the South African Police Services Forensic Science Laboratories, attempt to deal with these issues. Before these legal and scientific issues are addressed, it is essential to understand the rudiments of DNA (deoxyribonucleic acid) analysis. Forensic DNA analysis involves the intersection of molecular biology, genetics and statistical analysis. In order to assess the usefulness and the limitations of DNA evidence, it is important to be familiar with some of the basic principles underlying these disciplines.

In the investigation of crime use is made of DNA profiles from two sources: from the human body itself, and from small samples of human bodily material.

**Biological principles of DNA**

The relevance of DNA to forensic purposes is that, with the exception of identical twins, every person's DNA is unique. DNA can be extracted from blood, semen, bone tissue, saliva, hair roots and teeth. It is the genetic material passed from parent to child. Among human beings 99.99% of DNA sequences are identical. This shared DNA creates human characteristics that are similar in all people, such as having two hands and ten toes. The remaining .01% of DNA that is not shared is different in every individual, except in identical twins.

---

DNA (Deoxyribonucleic Acid) is a double-stranded molecule that contains the genetic code composed of 46 rod-shaped chromosomes, 23 of which are inherited from the mother, and 23 from the father.

The human body is made up of cells. Nearly all these cells (except, for example, mature red blood cells) contain a nucleus and a cytoplasm. The DNA is the same within and throughout a human being's cells. Scientists have developed techniques to identify the variations within an individual's sequencing. These techniques form the basis for DNA profiling.

The DNA in each human cell nucleus is around 3 billion nucleotides long. Each of the bases is linked not only with its neighbours in the long strands, but also across to another base in a parallel strand, creating a ladder-like structure. In the DNA molecule this ladder is twisted into the famous 'double helix'. The three billion base pairs do not form a single continuous chain, but coil up into separate sections, called chromosomes.

The nucleus contains two important structures, chromosomes and nucleoli (small round bodies of protein in a cell nucleus). In each human cell there are 46 chromosomes with approximately 100 000 genes carried in these chromosomes. The chromosomes are arranged in 23 pairs and one chromosome per pair is inherited from each parent. The 23rd pair is different from the others, as it determines an individual's sex. An offspring always receives an X chromosome from its mother but may receive either an X or a Y from its father. Individuals with XX in the 23rd chromosome are female while those with XY are male. The chromosomes combine to form a genome or genetic code.

A gene is found at a particular site or locus on a particular chromosome. Each distinctive sequence that may be found is an allele.

---

4 Ibid.
6 GT Duncan and ML Tracey 'Serology and DNA Typing' in WG Eckert *Introduction to Forensic Sciences* 2ed (1997) 245.
7 Cherfas op cit (n3) 24.
8 Farkas op cit (n2) 12
9 Farkas op cit (n2) 49.
Allele: an allele is one of several alternative forms of gene. Alleles are inherited separately from each parent and for a given gene an individual may have two different alleles (heterozygosity) or the same allele (homozygosity).

Usually, individuals have two copies of each gene at a given locus — one from the father and one from the mother. At each locus examined by DNA tests a person typically has two alleles, one maternal and one paternal. This pair of alleles is called a genotype.

**GENOTYPE:** the genetic constitution of an organism as distinct from its expressed features or phenotype.

**DNA PROFILE:** the set of genotypes possessed by a person at two or more loci is a multi-locus genotype or DNA profile.

Structurally, DNA is a double helix — two strands of genetic material spiralled around each other. Each strand has a ‘backbone’ made of sugar and phosphate groups and a sequence of nitrogenous bases, also called nucleotides, attached to the sugar groups. A base is one of four chemicals (adenine, guanine, cytosine and thymine). The two strands of DNA are connected at each base. Although the bases within each strand can be in any order, the cross-links between strands are limited: base A will cross-link only to base T and base C will cross-link only to base G.

**NUCLEOTIDE:** the unit of DNA consisting of one of four bases — adenine (A), guanine (G), cytosine (C), or thymine (T), — attached to a phosphate-sugar group.

This means that the two strands in a molecule of DNA are complementary: knowing the sequence of one enables the other to be described. This is vital in allowing the DNA molecules to be copied, as is necessary every time a cell divides to form new tissue. The links between the two strands are hydrogen bonds. These are weak bonds which are very sensitive to the chemical conditions surrounding the molecule. When cells divide, small changes to the cell chemistry cause the hydrogen bonds to break, and the DNA molecules split into two component strands. Each half of the DNA molecule then picks up more bases to reassemble its complementary strand, thus making two complete versions of the whole molecule.

The chemical structure of everyone’s DNA is the same. The only difference between people is in the order of the base pairs. There are so many millions of base pairs in each person’s DNA that every person has a different sequence. Using just these sequences can identify every person. However, because there are so many millions of base pairs, the task would be very time-consuming. Instead, because of repeating patterns in DNA, scientists are able to use a shorter method. These patterns do not, however, give an individual ‘fingerprint,’ rather, they are able to determine whether two DNA samples are from the same person, related people, or un-related people. Scientists use a small number of sequences of DNA from the non-coding DNA (parts of the
DNA sequence which are not genes) that are known to vary a great deal among individuals, and analyse these to get a certain probability of a match. For ethical reasons non-coding DNA is used for forensic purposes. This precaution avoids victimisation and prejudice based on genetically determined characteristics and traits.

Non-coding DNA: The meaningful part of the message of DNA — the part that contains the code for proteins and therefore of genetically determined characteristics or traits is interspersed with letters that appear to serve no purpose. These meaningless letters are sometimes called junk DNA, but are nevertheless useful for forensic purposes.

DNA tests are useful for identification because DNA profiles are highly variable across different people, making it unlikely that two different people will have exactly the same profile. Of course, DNA profiles are not necessarily unique because different individuals may, by chance, have the same genotypes in one or more loci. The likelihood of such a chance similarity depends on both the rarity of the matching genotype at each locus and the number of loci examined.

In *People v Brown*, the Court of Appeal for the Fifth Appellate Circuit of California explained DNA biology in a way that is comprehensible to the non-scientist:

“We begin with some simplified biology. The genetics of a human cell can be compared to a library, the genome, composed of 46 books, each a single chromosome. The text contained in the books is written in DNA, the chemical language of genetics. The library is compiled by the owner’s parents, each of whom contributes 23 books, which are then matched up and arranged together in 23 paired sets inside the sacrosanct edifice of the nucleus. During embryonic development, the original library is copied millions of times so that each cell in the human body contains a copy of the entire library.

Twenty-two of the twenty-three paired sets of books are entitled “Chromosome 1” through “Chromosome 22”; externally, the two paired books of each set appear to be identical in size and shape. However, the twenty-third set, which contains information on gender, consists of one book entitled “Chromosome X” (given by the mother) and one book entitled either “Chromosome X” or “Chromosome Y” (given by the father and determining the sex of the library’s owner). The 22 sets comprising “Chromosome 1” through “Chromosome 22” address an enormous variety of topics describing the composition, appearance, and function of the owner’s body. In addition, they include a considerable amount of what appears to be nonsense. The two paired books of each set, one book from each parent, address identical topics, but may contain slightly different information on those topics. Thus, two paired books opened to the same page contain corresponding paragraphs, but the text within those corresponding paragraphs may vary between the books. For example, within the paragraph addressing eye colour, one book may describe blue eyes while the other book of the set may describe brown eyes.

The two corresponding, but potentially variant, paragraphs in the two paired

10 *People v Brown* 110 Cal Rptr 2 d 750 (Cal App 2001).
books are called alleles. If, for a particular topic (i.e. at a particular region or locus on the DNA), the allele from the mother is A and the corresponding allele from the father is B, the genotype at that locus is designated AB. The text of two corresponding alleles at any locus may be identical (a homozygous genotype, e.g., AA) or different (a heterozygous genotype, e.g., AB). Regardless, one person’s genetic text is, in general, extremely similar to another person’s; indeed, viewed in its vast entirety, the genetic text of one human library is 99.9 percent identical to all others. As a result, the text of most corresponding paragraphs varies only slightly among members of the population.

Certain alleles, however, have been found to contain highly variable text. For example, alleles are composed of highly variable text when they describe structures requiring enormous variability. Also, some alleles appear to contain gibberish that varies greatly, or repeated strings of text that vary not in text but in repeat number. These variants (polymorphisms) found at certain loci render each person’s library unique [Identical twins, however, share essentially identical DNA] and provide forensic scientists a method of differentiating between libraries (people) through the use of forensic techniques that rely on the large number of variant alleles possible at each variable locus. For example, the combined libraries of the human population may contain two variant alleles at a particular locus, three at another, nine at another, and so on. Since each person receives two alleles for each locus, the number of possible combinations is further increased.

When a sample of DNA — usually in the form of hair, blood, saliva, or semen — is left at the crime scene by a perpetrator, a forensic genetic analysis is conducted. First, DNA analysts create a genetic ‘profile’ or ‘type’ of the perpetrator’s DNA by determining which variants or alleles exist at several variable loci. Second, the defendant’s DNA is analyzed in exactly the same manner to create a profile for comparison with the perpetrator’s profile. If the defendant’s DNA produces a different profile than the perpetrator’s, even by only one allele, the defendant could not have been the source of the crime scene DNA, and he or she is absolutely exonerated. If, on the other hand, the defendant’s DNA produces exactly the same genetic profile, the defendant could have been the source of the perpetrator’s DNA — but so could any other person with the same genetic profile. Third, when the perpetrator’s and defendant’s profiles are found to match, the statistical significance of the match must be explained in terms of the rarity or commonness of that profile within a particular population, that is, the number of people within a population expected to possess that particular genetic profile, or, put another way, the probability that a randomly chosen person in that population possesses that particular genetic profile. Only then can the jury weigh the value of the profile match.’

**Step-by-step overview of DNA profiling currently**¹¹ used by SAPS forensic science laboratories (FSL)

The concepts discussed below apply to any forensic DNA typing performed in forensic DNA laboratories internationally. The SAPS FSL form
part of this international discourse and adhere to these internationally accepted standards and procedures.

The PCR or Polymerase Chain Reaction\(^2\)

The PCR process produces millions of exact copies of the DNA at the specific locus to be analysed. This amplification of the initial DNA results in sufficient quantities for analysis. The PCR technique simulates the process which takes place when DNA is copied prior to the division of cells in the body. Each cell in the body contains a full copy of an individual’s DNA.

The Polymerase Chain Reaction (PCR) is an amplification technique used by the SAPS FSL as well as internationally. Previously it was used only when the DNA sample was small or when the sample was degraded by chemical impurities or damaged by environmental conditions. Nowadays PCR is used irrespective of whether the sample is limited or degraded or not.

The PCR process involves a series of cycles in which the DNA, combined with a cocktail of reagents needed to duplicate the target STR sequence (locus), is subjected to a range of discrete temperatures. The PCR process can be divided into three different steps depending on the temperatures. These are enzyme activation and separation of the original DNA into separate strands; the binding or annealing of the primers (synthesized DNA fragments which border on the area to be duplicated) to the separated strands, which assists in the targeting of the particular locus; and the extension of the primers using the original DNA strand to make a complementary copy of the original DNA which was present in the starting material. This three step cycle is then repeated in order to generate multiple copies of the specific areas on the original DNA.

Step 1: During this step the enzyme which is responsible for copying the original DNA (also called template DNA) is activated. The enzyme is supplied in an inactive form and needs to be activated. The high temperature conditions used in the process denature the DNA, that is it separates the complementary DNA strands to make the base sequence more accessible for the enzyme to initiate the copying process. The double helix of the DNA unwinds, leaving single-stranded DNA.

Step 2: The primers are annealed to the DNA. This enables the primers in the cocktail to anneal or bind to a specific segment of the single-stranded sample DNA. A primer is a short section of synthetically made DNA with a sequence complementary to the conserved region

---

\(^2\) This section is based on information provided by Senior Superintendent Arnold Greyling of the SAPS Forensic Science Laboratories. This usually forms the basis of the DNA expert’s testimony in court. Blocked sections and sections italicised indicate additional notes to this section. Any errors remain mine.
flanking the area to be amplified. The primers act as anchor points for the enzyme which demarcates the locus which needs to be copied, and allows for the enzyme to start the copying process.

Denaturing: a process by which the hydrogen bonds on the original double-stranded DNA are broken, leaving a single strand of DNA, the bases of which are available for copying into double stranded DNA again through hydrogen bonding.

Step 3: The DNA is extended by the enzyme with the primer as starting point. This allows for the various building blocks (the bases) to be added, thus extending the primer DNA and creating an exact copy of the original DNA at the locus. This process is repeated 28 to 30 times so that the specific section of DNA, specified and demarcated by the primers, is built to create millions of copies of the original DNA. With each cycle, the new copies generated act as additional templates for the next cycle. The increase of the copies is therefore exponential.

After PCR the DNA fragments produced are separated by fragment length using a technique called gel electrophoresis. The characteristics of the gel are such that the DNA fragments are separated according to length in the applied electrical.

The PCR thermal cycler

The above-mentioned PCR process takes place in an instrument called a thermal cycler. It is imperative that these instruments are in proper working order at all times. Therefore the correct functioning of these instruments should be verified frequently, at least annually. Negative and positive controls are used as a standard with the typing of all forensic samples, which also verifies the correct functioning of these instruments each time an analysis is done.

**What will happen if the thermal cycler does not function properly?**

Depending on which of the temperatures or times are affected, the impact will be very limited. The reagents used have been tested for such possible scenarios, and in most instances temperature differences varying with even two or three degrees Celsius will have very little noticeable impact on the DNA profiles or the correctness of the results. Slight time variations will also not have a significant impact. It is only when significant time as well as temperature differences are experienced during the PCR process that the specificity of the reaction and the success of the outcomes are affected. Problems with the thermal cycler will in most instances at worst lead to the unsuccessful amplification of a sample and in extreme situations to the false exclusion of a suspect. It is highly unlikely that thermal cycler problems could lead to
false inclusions of a suspect due to the nature of how temperature and time affect the DNA profile outcome.

There has already been a South African case, *S v Motloutsi*[^13], which challenged DNA evidence and temperature malfunctions were identified in the thermal cycler. However this had no impact on the correctness of the DNA result. Regular maintenance of instrumentation ensures that such incidents do not occur. In addition results are usually duplicated and therefore it is unlikely that the problem will recur.

The concept of Short Tandem Repeats (STRs)

The current DNA technique of choice internationally is called STR typing and refers to the specific type of DNA sequences targeted during the PCR process. STRs or Short Tandem Repeats is a sequence of bases which is repeated numerous times and is attached one after the other, in tandem, and therefore is referred to as a tandem repeat.

STRs at a specific locus differ between individuals according to the number of repeats. The number of repeats in turn is used to name the alleles, for example five repeats of the sequence four base sequence ATCG, that is ATCG ATCG ATCG ATCG ATCG will therefore be called allele 5. It is these alleles for each of the loci which are indicated in the table in SAPS FSL affidavits. Remember that one allele at each of the two loci typed is inherited from each of the parents. Therefore the results in the report received from the laboratory would, for example, indicate that at the STR locus D3S1358 the suspect’s sample generated an allele 5:5, if the same allele was inherited from both parents; and for example, if a 5 and a 7 allele were inherited from the respective parents the allele generated would be 5.

The hypervariability in the length of these tandem repeats in different individuals make them ideal for forensic purposes. The DNA flanking tandem repeat regions are conserved, meaning that a specific sequence of the DNA bases is the same amongst individuals in the population. This allows for the use of the PCR (Polymerase Chain Reaction) process in amplifying the tandem repeat region producing DNA fragments of varying length.

[^13]: In *S v Motloutsi* CPD CC 48/95 (unreported) at 36-42, maintenance of the thermocycler used in typing evidentiary material was challenged by the defence when it was discovered that components essential to ensuring consistent temperature changes were unavailable at the time of the testing. Furthermore, once the machine was tested, it was discovered that the 4°C storage facility was not functional.
Advantages of STR systems

DNA analysis based on STR analysis has the advantage of being more sensitive than previous forensic DNA analysis methods (for example HLA DQA1 and PM) since only a very small amount of template DNA is needed to obtain a result. The use of automated DNA sequencers also increases sensitivity. STR systems are more likely to be successful on old and poorly stored specimens that contain only degraded DNA when compared to older methods. The reasons for this are that the size of the STR fragments (100–400) base pairs are relatively small and the integrity of smaller lengths of DNA are more likely to be maintained in older samples. A base pair is two complementary nucleotides bonded together at matching bases (A and T or C and G) along the double helix of the DNA molecule. Base pair numbers are used to describe the location of an allele on the DNA strand.

The STR fragments can be accurately sized and analysed by automated sequencers using fragment analysis software. The use of the sequencers and software increases the amount of controls and internal checks. The chances of human error in analysis and interpretation are therefore kept to the absolute minimum. All data generated by means of STR technology can therefore be digitally archived for later reference, re-analysis or statistical analysis. Using one STR locus can already yield a high level of genetic discrimination. When a second or third locus is used in a multiplex PCR system, the levels of discrimination become very high. Therefore, the more STR loci included in the analysis, the more powerful the discriminating power and the more individual the STR profile.

STR systems in use at the SAPS forensic science laboratories

The STR system being used most frequently by the South African Police Services Forensic Science Laboratories is the AmpFISTR Profiler Plus™ PCR Amplification kit. This is a ten locus STR system which was developed and is supplied by the Applied Biosystems Corporation (United States of America).

In Australian and in several United States courts the defence have argued that evidence derived from this test kit was inadmissible because the manufacturer (Applied Biosystems) refused to disclose the primer sequences used in the kits (proprietary information) and the develop-


ment validation data generated by the manufacturer during validation of the kits. This defence was also subsequently raised in United States cases in efforts to have DNA evidence excluded.\textsuperscript{16} However, in \textit{R v Karger}, the court held that the evidence was admissible on the basis that the Profiler Plus test kit is recognised and accepted by the relevant (forensic) scientific community as reliable.

Recently, the Validation Homepage\textsuperscript{17} has been created by the US National Institute for Standards (NIST) with the purpose of providing details on studies conducted, a description of samples tested and the number of samples run as part of the study in order to assist current and future validation efforts by forensic DNA laboratories. The validation of this kit for its application to forensic samples was performed to meet the international requirements as recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines for validation. Every positive control typed correctly by the kit validates the kit since it proves the reliable use of the kit by the forensic labo-

\textsuperscript{16} In \textit{State v Pfenning}, No 57-4-96 (Vt. Dist. Ct. Apr. 26, 2000), the trial court excluded the results of both Perkin-Elmer (PE) and Promega kits because the kit makers had not released enough information for the court to be sure of the kit's reliability under \textit{Daubert v Merrell Dow Pharmaceuticals, Inc}, 509 U.S. 579 (1993). See also \textit{State v Dishmon}, Nos. 99047345, 99069306, 99079650, slip op. at 13 (Minn. Dist. Ct. Mar. 3, 2000) (noting that in resisting the defence subpoena, (“PE Biosystems claimed trade secrets, overwhelming expense and effort[,] and [] reading between the lines[,] claims that the material is not needed here”); In \textit{People v Cavin}, No 004-4395-NY (Mich. Cir. Ct. Oct. 18, 2000), a Michigan trial court case, PE agreed to reveal its primer sequences provided all parties involved agreed to a protective order. One Colorado trial court noted that ‘PE has resisted releasing its developmental data claiming that the data was unavailable, that it had never been systematically recorded that it was scattered throughout various departments at the company and [that] its collection at this time would be unduly burdensome’. \textit{People v Schreck}, No 98CR2475, slip op. at 13 (Colo. Dist. Ct. 2000) (granting the defendant's motion to bar DNA evidence derived from multiplex kits), available at: \url{http://www.scientific.org/distribution/archive/Schreck.doc} (on file with the \textit{Duke Law Journal}). The Colorado Supreme Court later vacated the decision. \textit{People v Schreck}, 22 P3d 68, 81 (Colo. 2001). In at least two cases trial courts have held kit results admissible even though the kit maker refused to comply with a defence subpoena for data relating to the kit. The court in \textit{Dishmon} held that the denial of discovery was not significant to admissibility of the kit because (1) the State could make the required showing of admissibility of the results without the discovery and (2) the defendant had access to other information and thus did not suffer any prejudice that would violate his right to a fair trial. \textit{Dishmon}, Nos. 99047345, 99069306, 99079650, slip op. at 15. The \textit{Dishmon} court makes questionable assumptions as to the sufficiency of the evidence to which the defendant was given access. Another court found a subpoena unenforceable because the burden on the kit maker substantially outweighed the alleged demonstrated need for the information. \textit{People v Bertsch}, No. 9F07295 (Cal Super Ct Oct 20, 1999). See \url{http://www.denverda.org/html/website/denver_da/dna_resources.html}, accessed on 2005/09/10 where courts rejected the defence challenges.

\textsuperscript{17} \url{http://www.cstl.nist.gov/biotech/strbase/validation.htm} \url{http://www.cstl.nist.gov/biotech/strbase/validation.htm}, accessed on 14/10/2004.
ratory with an international traceable standard. The same is true for the use of the Internal lane standard and allelic ladders used during electrophoresis and analysis of the STR results.

The AmpFlSTR Profiler Plus™ kit looks at ten different places (loci) on an individual’s DNA to generate a so-called STR profile. These areas are called D3S1358; vWA; FGA; D8S1179; D21S11; D18S51; D5S818; D13S317 and D7S80. The tenth area, named Amelogenin, indicates the sex of the individual from whom the exhibit or reference sample originates.

Laboratory controls

Positive controls

A positive control is included with the kit to test whether the reagents were set up correctly, the PCR process and electrophoresis process were performed properly, and the analysis of the results with the software was performed correctly by the laboratory. For example — the kit of reagents will contain DNA 9947A which is a Standard Reference Material (SRM) and has an internationally known and accepted STR profile.

If the correct genotypes are obtained with the positive controls, it is an indication that the process has worked properly and that all the parameters and conditions needed in the PCR and analysis process were correct and are traceable to an international standard.

Negative controls

The negative control consists of all of the components of the PCR set-up with the exception of sample DNA and shows if there was a possibility of contamination. These controls undergo precisely the same process as the unknown samples. If the negative control shows no amplification product, it indicates that no contamination of any of the reagents used in the PCR process took place.

Duplication or repetition of the laboratory tests

Another safeguard to minimize the risk of error by the laboratory is to repeat or duplicate the process — from DNA extraction to interpretation. If a laboratory error in typing of the samples is suspected, the best way of verifying the results is to have the samples retested by a different laboratory. For this reason, in accordance with international guidelines, and wherever the amount of sample permits, the SAPS FSL retains a portion of the original sample for retesting by an independent
laboratory. A known sample for internal proficiency testing is included with every run (See also Quality Control infra).

After PCR the DNA fragments produced are separated with regard to fragment length by using a technique called gel electrophoresis or capillary electrophoresis. Different types of instruments can be used to perform the electrophoresis process but in general, all can be divided into two groups, namely slab gel and capillary instruments, based on the separation technique. After the separation of the fragments the instrument sizes the fragments by means of software and subsequently assigns allele names to peaks on what is called an electropherogram. An electropherogram is a software generated graph which represents the DNA profile, but it is the allele results which are used by the forensic analyst to interpret the results within the context of the case.

**Gel electrophoresis**

Gel Electrophoresis: a technique used to separate molecules such as DNA fragments or proteins; in forensic uses of DNA tests, electric current is passed through a gel, and the fragments of DNA are separated by size; smaller fragments will migrate further than larger pieces.

The STR profile is represented by what is called an electropherogram. The electropherogram shows the STR profile as a series of peaks where the peak height indicates the fluorescent intensity of the bands measured as relative fluorescent units (rfu) on the y-axis.

Each sample also contains an internal lane size standard (represented by red bands), which after it has been defined helps to accurately calculate the fragment size of each allele in the profile. In the above-mentioned electropherogram, the x-axis is displayed in fragment size measured as basepairs (amount of building blocks).

**How are Electropherograms produced?**

ProfilerPlus™ (An STR Reagent kit) uses ‘primers’ to identify the relevant STR-DNA segments and then ‘amplifies’ (replicates) these segments using a process called polymerase chain reaction (PCR). Each locus is ‘labelled’ with a coloured dye (either blue, yellow or green). The capillary electrophoresis instrumentation (e.g. the Applied Biosystems 310 and 3100 sequencer platforms) measures the length of the DNA segments by using an electrical current to impel them through a narrow capillary tube, wherein the shorter fragments move more quickly than the longer fragments. On slab gel electrophoresis instruments (e.g. the Applied Biosystems 377 sequencer platforms) the same principle applies, except that a capillary is not used, but a slab gel. Under laser light, the coloured dyes produce fluorescent light, signalling the presence of DNA. A computer-operated camera detects the

---

18 This information block includes more details about how STR testing is done, but is separated from the main text to avoid burdening readers who are not interested in the details.
light as the fragments migrate past a specific point of the capillary or gel. The ‘peaks’ on the electropherogram record these flashes of light. Based on the colour of the light, and the time it took the DNA to pass through the capillary, a series of computer programmes determine which alleles are present at each locus.

The position of the peaks on the graph (how far left or right) indicates how long it took the allele to pass through the capillary or gel, which indicates the length of the underlying DNA fragment. From this, the computer programme infers which allele is represented and generates the appropriate label. This label is called an allele designation. This allele designation is a standardised nomenclature used by all forensic DNA laboratories and provides for a standardized format with which different laboratories can communicate DNA profiles to one another.

The height of the peaks corresponds to the quantity of DNA present. The unit of measurement for peak heights is the RFU (relative fluorescent unit) which reflects the intensity of the fluorescent light detected by the computer-operated camera. Peaks representing alleles from the same person are expected to have roughly the same heights measured in RFUs throughout a given sample, although peak height imbalances occasionally occur. These imbalances do not impact on the correctness of the results.

Gel image (only generated on slab gel instrumentation, not on capillary electrophoresis systems)

**Internal Lane Standard and Allelic Ladder**

These are provided by the manufacturer of the PCR kit, and act as a control for the verification that the electrophoresis of the sample took place within acceptable parameters. The Internal Lane Standard (ILS) is run with every sample to ensure correct sizing of DNA fragments and assists in the accurate sizing of the STR fragments by the software. The ILS consists of fragments of known sizes and is also an international traceable standard used by all forensic DNA laboratories for this purpose. The Allelic Ladder contains all of the alleles most frequently encountered in the world populations for each of the STR loci which are typed by the forensic laboratory. The allelic ladder is also provided by the manufacturer of the STR kit and is used as a control to assist the software and forensic analyst in correctly designating the alleles in the STR profile. The unknown sample’s alleles are assigned by comparing the fragment sizes of the STR fragments in the sample (exhibit or reference sample), with the fragment size obtained for the alleles in the allelic ladder.

**STR result, analysis and interpretation**

The SAPS FSL performs several levels of analysis on the STR results generated during electrophoresis on the sequencer instruments before
finally accepting a DNA profile for interpretation within the context of a case. In general most forensic DNA laboratories use STR results only if these were analysed by means of two commercially available software applications respectively called:

Genescan® Analysis — this software uses the data collected by the sequencer during electrophoresis and performs the sizing of the STR fragments of the unknown samples relative to the first control, that is the Internal Lane Standard (ILS). It also provides the forensic analyst with the functionality to view the STR profile in more detail to establish the general quality of the profile.

Genotyper® software — this software uses the same data to compare the sized fragments of the unknown sample to the fragment sizes of the alleles in the second control, that is the allelic ladder. This process makes use of standard programmed parameters which are then applied to the data generated by the specific forensic laboratory. The final result is an electropherogram with allele labels. The allele labels for all the loci together provide the genotype profile (STR/DNA profile) for the individual or exhibit. It is this profile which is indicated in the table provided with the affidavit in a case.

Recently a new software application called GeneMapper ID® has been introduced for the same purpose. This is merely the incorporation of the two above-mentioned software application functionalities into a single software package. The underlying concepts are still exactly the same as those mentioned above, with some added features to assist the analyst in assessing possible STR profile quality issues.

The SAPS FSL takes this analysis process one step further by using a second software system, also called an Expert system, to verify the results generated by the Genotyper® software. During this phase of analysis a second analyst will verify whether the STR results generated during the first analysis is correct, and provide a second opinion on whether the quality of the STR profile is sufficient to ensure reliable interpretation. Note that whereas the first level analyst (using Genescan and Genotyper) does not have access to previous results obtained for the same sample or for that matter the other results obtained on samples within the same case except sometimes when in the same sample category (that is reference sample or evidence sample), the second level analyst has a unique view from both the sample and case perspective. This is due to the fact that the second level analyst has access to a history of previous results which may have been generated with the same sample, as well as other samples within the case. This second level analyst can therefore verify whether duplication of results has been possible, and can determine whether the STR profiles are of sufficient quality for the interpretation within the context of the case.
Three possible outcomes of the interpretation of the STR results are attainable:

Match — The STR (DNA) profile of one sample may match the DNA profile/s generated on other samples.

Non Match — The STR (DNA) profile of a sample does not match the profiles generated on other samples. For example: exclusion of a suspect after typing a crime scene sample, as well as the reference (control) sample of the suspect.

Inconclusive — this outcome will be stated if the laboratory was unable to state a clear match or non match based on the results for example in rape cases with multiple suspects, where a mixture of DNA profiles is generated.

Even after the interpretation of the STR results has been performed i.e. match, non match or inconclusive and an affidavit or letter to that effect has been generated — the interpretation is verified by a second independent analyst who ensures that the interpretation is scientifically objective and correct. Each STR result and its interpretation provided by the SAPS FSL has therefore been checked by at least two, but in most instances by at least three, analysts if not more. If any result could not be verified, the laboratory will issue a negative report.

Discovery issues

Forensic experts at the SAPS FSL are of the view that it is essential in discovery issues that there be open lines of communication between the laboratory, the lawyers concerned and the investigating officer.

Discovery by the SAPS FSL can assume two forms:

The SAPS FSL may provide the samples which were typed for a case to an independent laboratory or expert for re-testing in order to verify the interpretation provided by the FSL, or it may provide documentation to support the DNA typing process which was used to generate the results. These documents are intended specifically to assist a defence expert in verifying that the SAPS FSL did not commit errors of processing or interpretation.

Consultation with the SAPS FSL also raised the following issues:

The SAPS FSL stressed that excessive and unnecessary discovery requests should be avoided wherever possible. The SAPS FSL does furnish documentation for the verification of the DNA typing process, but only if the defence has obtained the services of a DNA expert. The
reason for this is that most of the material is highly technical and lay persons will experience problems in interpreting the data. Some of the information includes electropherograms which only scientists with experience of the technology will be able to evaluate. This could lead to delays in the court process. Another reason for curbing unnecessary discovery requests is that they impact severely on the SAPS FSL casework throughput. Demands for documentation on short notice require that the FSL allocate significant numbers of personnel to the process of compiling stacks of documentation from the process archives which, if not assessed by a scientist, will have limited value to the court. The officers of the court should therefore always clarify what the purpose for requesting discovery is, and whether there are specific issues which need to be clarified. Frequently counsel for the defence will request `all relevant documentation', with no reference to the specific aspect of the DNA evidence being challenged and without the assistance of an expert. This approach has a significantly negative impact on the caseload of the SAPS FSL.

To assist in discovery, the SAPS FSL has always made analysts available to discuss any possible issues of contention with defence experts. In many instances these problems can be easily resolved, as they are usually due to a lack of understanding the typing process rather than being due to laboratory mistakes. By having these pre-trial conferences scientists can identify the problems at issue without wasting time.

Problems and issues that have been encountered in DNA evidence

DNA profiling is widely accepted by the courts as a robust and reliable technique. Challenges tend to be on technicalities. Many of the problems encountered in the past have either been rectified, or were due to techniques no longer in use. *S v Maqhina* is a good example of the latter, as in 1997 the STR-technique replaced the HLA DQ-Alpha used previously.

Lawyers must be aware of the possible attacks that can be launched against DNA evidence and whether these arguments have any substance. There is a danger in using comparative case law because many precedents from other jurisdictions relate to procedures not used in South Africa.

Issues pertaining to sample quantity and quality

Several factors that may affect a DNA sample’s suitability for analysis. Questions that can be raised are the following:
Did the crime sample contain enough DNA to permit accurate analysis?

It has been argued that to be interpretable, the crime sample must contain enough DNA of sufficiently high molecular weight to allow isolation of longer DNA fragments as these are most susceptible to degradation. Samples of blood, semen, or other DNA sources may be too small to permit analysis. These arguments apply to older techniques such as Restriction Fragment Length Polymorphisms (RFLPs), but are not relevant to current technologies. An advantage of the STR system used by the SAPS Forensic Laboratories is that very small regions can be targeted. With PCR-based systems a single cell is, in theory, sufficient to generate a DNA profile. In practice it has been found that as little as approximately one billionth of a gram of material is sufficient to obtain a DNA profile. Less semen or saliva than blood is needed to obtain an equivalent type. This is because the concentration of sperm cells in semen, or epithelial cells in saliva, is higher than the concentration of white blood cells in blood. In general it is not so much the amount of DNA available as the quality that determines the possibility of generating a DNA profile.

Was the crime sample of sufficient quality to permit accurate analysis?

Exposure to heat, humidity, time, ultraviolet radiation and various chemical substances can degrade the DNA sample. With older techniques exposure to chemical or bacterial agents could alter DNA by interfering with the enzymes used in the testing process. This is no longer the case, as the high discriminatory power of the STR techniques will merely produce no results.

Rudin and Inman refer to numerous studies that have been conducted to determine the effects of these conditions:

‘An important outcome of these studies is the finding that these environmental factors will not change DNA from one type into another; in other words, an HLA DQA7 type 1,1 will not change into a 1,2, nor will an STR type change from a 5,9 to a 6,8. Rather, the degradation changes the DNA from a sample that can be typed into a sample that gives no type at all. This is an important part of the validation of any genetic typing system because it means that the biological component of the system will not produce false positive results.’

In other words, because one profile cannot be changed into another, there is no danger that environmental degradation will produce a complete DNA pattern that would include someone who is not the donor.

Rudin and Inman \(^{19}\) refer to numerous studies that have been conducted to determine the effects of these conditions:

‘An important outcome of these studies is the finding that these environmental factors will not change DNA from one type into another; in other words, an HLA DQA7 type 1,1 will not change into a 1,2, nor will an STR type change from a 5,9 to a 6,8. Rather, the degradation changes the DNA from a sample that can be typed into a sample that gives no type at all. This is an important part of the validation of any genetic typing system because it means that the biological component of the system will not produce false positive results.’

In other words, because one profile cannot be changed into another, there is no danger that environmental degradation will produce a complete DNA pattern that would include someone who is not the donor.

\(^{19}\) N Rudin and K Inman *Forensic DNA Analysis* 2ed (2002).
of the sample. Degradation can limit the usefulness of DNA typing, but does not invalidate it.

The SAPS FSL recently introduced new technology for the improved evaluation of sample quality and more accurate determination of sample DNA quantities. This technology is called Real Time PCR, and allows the SAPS FSL to determine the quantity and quality obtained from evidence samples; the presence of male DNA in rape case evidence; and the possible presence of mixtures of DNA in evidence samples.

This technology will allow the SAPS FSL to apply resources to cases where the chances of successfully generating DNA profiles are most likely.

Mixtures contained in DNA samples

The sample can often be identified as coming from a single perpetrator, or a single victim, or from both. A crime sample of blood or semen may include DNA from multiple sources, as when more than one person has contributed to the sample. Male and female DNA extracted from such a sample can be distinguished, as can same-sex DNA where the alternative sources are known and available for testing (such as a rape victim’s husband). The presence of multiple, same-sex samples from unknown sources previously caused additional complications, but with the multiplex STR system this is no longer a problem. The amelogenin locus (the gene for tooth pulp), shows a length variation between the sexes, and it is therefore possible to identify DNA profiles originating from more than one individual with ease.

Issues pertaining to laboratory performance

DNA profiling is valid and reliable but confidence in a particular result depends on the quality control and quality assurance procedures in the laboratory, and on the handling. Quality control refers to the nature of the samples that are being compared.

Quality control and assurance

Quality assurance refers to monitoring, verifying, and documenting laboratory performance. Note that a quality assurance programme, and not accreditation, helps to demonstrate that a laboratory is meeting its quality control objectives and thus justifies confidence in the quality of its product.

---

20 For general descriptions of quality assurance programmes in the United States, see National Research Council (NRC) at ch 3 ‘Ensuring High Standards of Laboratory Performance’; NRC II, at ch 4 (1996).
Professional bodies within forensic science have detailed procedures for quality assurance. Guidelines have been prepared by two FBI-appointed groups, the Scientific Working Group on DNA Analysis Methods (SWGDAM)\textsuperscript{21} and the DNA Advisory Board (DAB).\textsuperscript{22} The American Society of Crime Laboratory Directors-Laboratory Accreditation Board (ASCLD-LAB) accredits forensic laboratories.\textsuperscript{23} Since 1998, ASCLD-LAB has accredited laboratories in Australia, New Zealand and Hong Kong as well as laboratories in the United States and Canada. Note that the ASCLD-LAB accreditation programme does not allow laboratories to obtain accreditation only for a particular service — for example DNA typing — a laboratory seeking accreditation must qualify for the full range of services it offers. A laboratory providing, for example, services related to ballistics examinations, questioned document examinations and drug analysis will therefore have to be accredited within all these fields. This constraint has slowed and prevented many forensic DNA labs from seeking accreditation.\textsuperscript{24}

In \textit{S v Maqhina}\textsuperscript{25} much was made of the fact that the SAPS Forensic Science Laboratory is not an accredited laboratory. In fact, there are no laboratories in South Africa that are accredited for the forensic DNA process or for that matter as forensic DNA laboratories. Most international standards for laboratories (for example ISO 17025; DNA Advisory Board (USA); SWGDAM) place the emphasis on a requirement of an internal Quality System at a laboratory as an indicator that good laboratory practices (GLP) are in place, rather than accreditation. The South African National Accreditation System (SANAS) has to date drafted but not finalised standards for laboratories which wish to be accredited in this field.\textsuperscript{26} SANAS is therefore not in a position as yet

---

\textsuperscript{21} Scientific Working Group on DNA Analysis Methods, `Guidelines for a quality assurance program for DNA analysis,' 22 [hereinafter SWGDAM Guidelines].


\textsuperscript{23} American Society of Crime Laboratory Directors-Laboratory Accreditation Manual Jan 1997.

\textsuperscript{24} As an interim solution, the National Forensic Science Technology Centre (NFSTC) has an agreement with ASCLD-LAB to perform certification audits on DNA sections of laboratories for compliance with DAB and ASCLD-LAB standards; this service is available to private sector DNA laboratories as well as government laboratories. This is a solution that may be investigated by the SAPS Forensic Laboratory.

\textsuperscript{25} \textit{S v Maqhina} 2001(1) SACR 241 (T).

\textsuperscript{26} These standards were generated with the input of a Specialist Technical Committee consisting of representatives from various paternity laboratories, tertiary institutions and the SAPS FSL. The quality assurance standards and guidelines are based on similar standards by the DNA Advisory Board (USA).
to accredit the SAPS FSL in all fields. Due to the cost involved and the variety of fields of expertise, the SAPS prefers not to be accredited before all aspects can be accredited at the same time. Accreditation is not the solution which the legal system requires as proof of reliability. It is for these reasons that the SAPS FSL biology unit refers to its own quality system in court, and not to national standards which are not yet fully in place. The SAPS FSL does not intend to be accredited in the near future, even if these standards were to be finalised, since some of the services provided by the laboratory have no national or international accreditation guidelines or standards available as yet.

What must be realised is that accreditation in itself does not assure continuous compliance, nor does it guarantee a zero error rate. What is perhaps more important from a quality assurance point of view is to establish what quality control measures are in place and whether they are adequate to ensure valid and reliable results.

The quality assurance guidelines promulgated by SWGDAM, the DAB, ASCLD-LAB and ENSFI call for laboratories to document laboratory organization and management, personnel qualifications and training, facilities, evidence control procedures, validation of methods and procedures, analytical procedures, equipment calibration and maintenance, standards for case documentation and report writing, procedures for reviewing case files and testimony, proficiency testing, corrective actions, audits, safety programmes, and review of subcontractors. Maintaining extensive documentation and records can contribute to verifying the correctness of results obtained in any particular case. However, errors in analysis or interpretation might occur as a result of a deviation from an established procedure, analyst misjudgement, or an accident. To prevent these errors in analysis and interpretation, the SAPS FSL follows a policy of case review, validation, proficiency testing and good practice procedures in connection with the identification, collection and handling of samples.

**Case-review**

Case-review procedures within the SAPS Forensic Laboratories are designed to detect errors before a report is issued. In addition to the laboratory maintaining a strong quality assurance programme, it also follows a case-by-case review process — whereby a second analyst will always review the result. Part of the quality control exercised within the biology unit is by the STRiker analyst and the STRgazer analyst. The STRgazer analyst's role is objective as he or she is responsible for the quality control of DNA profiles at sample level. The sample results are evaluated independently from the case, and the STRgazer analyst evaluates the quality of the STR profiles without interpretation of matches or non matches. The STRiker analyst, in turn, is responsible for the quality control of the
DNA results at the case level and provides a technical review of the result interpretation done by the reporting officer (the analyst responsible for the particular case). It is a primary duty of the reporting officer to give an unbiased scientific opinion. However, to guard against the possibility of conscious or unconscious bias, the role that the STRiker analyst plays is of cardinal importance as a quality control measure who ensure that the scientifically objective reporting of facts takes place.

Validation

The validation of procedures is central to quality assurance. Internal validation involves the verification by a laboratory that it can reliably perform an established procedure which has already undergone developmental validation. Before adopting a new procedure, the SAPS FSL verifies its ability to use it. Validation builds on the accumulated body of knowledge and experience. Some aspects of validation testing need be repeated to verify that previously established principles apply. The SAPS FSL will therefore not necessarily repeat studies which were performed by the manufacturer of the kits during the development stage (developmental validation). It will perform internal validation studies to confirm that the product functions within acceptable parameters. It should be noted that every positive control typed correctly by the SAPS FSL during casework acts as verification that the technology was applied correctly to the samples because these are internationally traceable standard materials.

Proficiency testing

Proficiency testing in forensic genetic testing is designed to ascertain whether an analyst can correctly determine genetic types in a sample, the origin of which is unknown to the analyst, but is known to a tester. Proficiency is demonstrated by correct genetic typing in repeated trials. Proficiency tests also require laboratories to report random-match probabilities to determine if proper calculations are being made. The DAB recommends that every analyst should regularly undergo external, open proficiency testing and that the laboratory should take ‘corrective action whenever proficiency testing discrepancies [or] casework errors are detected.’

---

27 See NRC Report II (1996): ‘... case files be reviewed by a qualified second analyst before a report is released and ... laboratory procedures should be designed with safeguards to detect bias and to identify cases of true ambiguity’.

28 Standard 13.1 specifies that these tests are to be performed at least every 180 days. DAB Standards, op cit (n22) 16. SWGDAM recommended two open proficiency tests per year per analyst. SWGDAM Guidelines, op cit (n21).

29 DAB Standards op cit (n22) 17 (standard 14.1).
Identification, collection and handling of samples

Good practice in recovering and handling of DNA crime samples can maximise the forensic capabilities of DNA profiling. SAPS subscribes to systems as provided for in ISO/IEC 17025 as well as those contained in SANAS R42-01. Failure by law enforcement personnel to identify and collect appropriate DNA evidence from the crime scene can disadvantage the investigation and prosecution of serious crimes. SAPS FSL personnel and the whole STR typing process are regularly subjected to both internal and external proficiency testing.

Interpretation of laboratory results

The ultimate purpose of DNA typing is to test the hypothesis that a particular person is the source of an item of biological evidence. An attempt is made to ascertain whether two samples, evidence and reference, share a common source. The evidence sample, which can consist of a biological fluid or tissue and the reference, usually a blood sample, are subjected to DNA tests as described above. Upon completion, the analyst is able to render a determination as to the possible source or sources of the samples.

Exclusions, inclusions, and inconclusive results

Table 1: Possible Outcomes of DNA Matching

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Description</th>
<th>Explanation/Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusion/Negative result</td>
<td>Profiles are different — no DNA match</td>
<td>This is evidence that because the profiles are different, they accordingly may have originated from different sources.</td>
</tr>
<tr>
<td>Inconclusive/Null result</td>
<td>Profile comparison not possible</td>
<td>Here a comparison is not possible based on the results of the tests and after review by both the Reporting officer and the STRiker analyst. This would most frequently occur in the case of mixtures.</td>
</tr>
<tr>
<td>Conclusive/Positive result</td>
<td>No differences were observed between the samples (Any difference will immediately become an exclusion)</td>
<td>The samples present a genetic concordance of several polymorphic DNA loci (such as in STR) and can be evidence that the two profiles share a common source.30</td>
</tr>
</tbody>
</table>

30 See below how alternative hypotheses for a positive result are dealt with. In the first instance, it is important to know that the current technology used for DNA sampling, the AMPFISTR™ Profiler Plus PCR Amplification kit has a very high level of discrimination. Apart from Amelogenin which can determine sex and the nine other loci that are tested look at non-coding DNA.
Once the results of the DNA typing procedures of a common source can be inferred between the evidence sample and the suspect sample, the question arises: What is the strength of this result? The answer will depend on the probability of finding this profile if the suspect were the true source, compared to the probability of finding this profile if someone other than the suspect were the true donor. If many individuals share the profile, then the strength of the match is minimal, because a reasonable chance exists that someone else from the population has the same type. If only a low probability exists that the types found are from someone other than the suspect, then the inference of a common source is strong. The loci tested by the kit are such that if a match is found, the chances are one in one billion (depending on how common the DNA profile is amongst the population) that someone other than the suspect was the donor. However, before frequency estimate calculations can be made, it is important that all alternative hypotheses for finding a match must be excluded.

**Alternative hypotheses**

If the accused is the source of DNA found at a crime scene, then a DNA sample from the accused and the forensic sample should have the same profile. The forensic scientist will then report that the sample of DNA from the crime scene and a sample from the accused have the same genotype. The question of importance is whether this proves that the accused is the source of the forensic sample? It is possible that other explanations could account for the matching profiles, such as laboratory error. This could arise from mistakes in labelling or handling samples, or from cross-contamination. The 1992 NRC report in the United States cautioned, ‘[e]rrors happen, even in the best laboratories, (even in accredited laboratories) and even when the analyst is certain that every precaution against error was taken.’

Another possibility is that the laboratory analysis is correct, but the forensic sample came from another individual. In general, the true

---

31 It has been argued that the fact that the accused is the source does not necessarily mean that the accused is guilty of the offence charged. Aside from issues which bear on intent or knowledge and that have nothing to do with DNA, there remains, for instance, the possibility that the two samples match because someone framed the accused by putting a sample of defendant’s DNA at the crime scene or in the container of DNA thought to have come from the crime scene. See generally United States v. Chischilly, 30 F3d 1144 (9th Cir. 1994) (dicta on ‘source probability’); J J Koehler ‘DNA matches and statistics: important questions, surprising answers’ (1993) 76 Judicature 222.

32 NRC I report at 89.
source might be a close relative of the accused or an unrelated person who just happens to have the same profile as the accused. The former hypothesis refers to kinship, and the latter to coincidence. For the inference to be drawn that the accused is the source of the crime scene DNA these alternatives must be discounted. If they can be then only the hypothesis of identity remains. The considerations that affect the chances of a reported match when the accused is not the source of the trace evidence are briefly discussed below.

Dealing with laboratory error

Although many experts would concede that even with rigorous protocols, the chance of a laboratory error exceeds that of a coincidental match, quantifying the former probability is a formidable task. First, the SAPS FSL has quality assurance systems in place to prevent error. All forensic samples are divided in two, so that if error is suspected, the defence has the opportunity to have the sample retested. The FSL policy is based on the principles set out in the NRC report II in a section entitled ‘Should an Error Rate be Included in Calculations?’

The report lists four reasons why laboratory error should not be combined with random match probability calculations. In the first instance the relevant statistic is not a general error rate for the testing laboratory, or laboratories in general, but whether the laboratory has committed an error in this particular case. Risk of error in a particular case, however, depends on many variables, and no simple equation exists to translate them into a probability statistic. Secondly, a testing laboratory would have to undergo an unrealistically large number of proficiency tests to allow the estimation of a statistically valid probability of error. Thirdly, although it might be possible to generate a probability statistic by pooling data from proficiency tests of several laboratories, so producing an ‘industry-wide’ error rate, this statistic would unfairly penalise better laboratories. Fourthly, an error rate estimated by a laboratory’s historical performance on proficiency tests will almost certainly be too high, since errors will be investigated and

35 A close relative, for these purposes, would be a brother, uncle, nephew. For relationships more distant than second cousins, the probability of a chance match is nearly as small as for persons of the same ethnic subgroup. For an instance of the ‘evil twin’ defence, see Hunter v. Harrison No. 71723, 1997 WL 578917 (Ohio Ct.App 18 Sept 1997) (unpublished paternity case).


37 NRC II Report at 85-7.

38 NRC II Report at 85-6.

39 NRC II Report at 86.
corrected after they are discovered, and thus errors committed in the past are not likely to recur. The committee concluded:

‘[W]e believe that a calculation that combines error rates with match probabilities is inappropriate. The risk of error rates with match probabilities is inappropriate. The risk of error is properly considered case by case, taking into account the record of the laboratory performing the tests, the extent of redundancy, and the overall quality of the results.’

The report went further and stated that an accused’s ‘best insurance’ against the possibility of a false match due to laboratory error is the opportunity to have testing repeated at another facility.  

Kinship among suspect population

With enough genetic markers, all individuals, except for identical twins, should be distinguishable. This, however, was not always possible with the limited number of loci previously used in forensic testing. Close relatives have more genes in common than unrelated individuals, and consequently various procedures have been proposed for dealing with the possibility that the true source of the forensic DNA is not the accused, but a close relative. Recommendation 4.4 of the 1996 NRC report which the SAPS Forensic Laboratory follows, reads:

‘If possible contributors of the evidence sample include relatives of the suspect, DNA profiles of those relatives should be obtained. If these profiles cannot be obtained, the probability of finding the evidence profile in those relatives should be calculated with [specified formulae].’

Often the investigation, including additional DNA testing, can be extended to all known relatives but this is not always feasible, and there is always the chance that unknown relatives exist in the suspect population. When that population is very large, however, the presence of a few relatives will have little impact on the probability that a suspect drawn at random from that population will have the incriminating genotype. Furthermore, it has been suggested that the effect of relatedness is of practical importance only for very close relatives, such as siblings. Formulae are available for computing the probability that any person with a specified degree of kinship with the suspect also possesses the

---

58 See People v Reeves 109 Cal Rptr 2d 728 (Cal App 2001) 751-52.
59 See for example, B S Weir ‘Discussion of Inference in Forensic Identification’ 158 Royal Stat. Soc’y Ser A 49 (1995) (‘the chance that two unrelated individuals in a population share the same 16-allele [VNTR] profile is vanishingly small, and even for full sibs the chance is only 1 in very many thousands.’)
60 NRC II op cit note 20 at 113.
incriminating genotype.\textsuperscript{42} For example, the probability that an untested brother (or sister) would match at four loci (with alleles that each occur in 5\% of the population) is about 0.006; the probability that an aunt (or uncle) would match is about 0.000000.\textsuperscript{43}

Coincidence with rival hypothesis

Another rival hypothesis is coincidence, where the accused is not the source of the crime scene DNA, but happens to have the same genotype as an unrelated individual who is the true source. Various procedures for assessing the plausibility of this hypothesis are available. In principle, one could test all conceivable suspects. If everyone except the accused has a non-matching profile, then the conclusion that the accused is the source is inescapable. Testing of the population of conceivable suspects is almost never feasible. The suspect population normally defies any enumeration, and in the typical crime where DNA evidence is found, the population of possible perpetrators is so huge that even if all its members could be listed, they could not all be tested.\textsuperscript{44}

The SAPS FSL Biology Unit has found that of the 19 000 DNA profiles on the database, not one has represented a match with another. The NRC II recommends that such additional information can be gleaned from the comparison of profiles within databases. In the overview to NRC II, it is stated at 34:

‘Additional information comes from comparison of profiles within the databases. An early study used FBI and Lifecodes data for Blacks, Whites, Southeast Hispanics and Southwest Hispanics. Among 8,628,360 pairs of profiles from within those databases, no four or five-locus matching profiles were found, and only one three-locus match was seen. A newer and more extensive analysis, compiling data from numerous TWGDAM sources, summarized a large number of profiles from White, Black and Hispanic databases. Of 58


\textsuperscript{43} The large discrepancy between two siblings on the one hand, and an uncle and a nephew on the other, reflects the fact that the siblings have far more shared genes. All their genes are inherited through the same two parents. In contrast, a nephew and an uncle inherit from two unrelated mothers, and so will have few maternal alleles in common. As for paternal alleles, the nephew inherits not from his uncle, but from his uncle’s brother, who shares by descent only about one-half of his alleles with the uncle.

\textsuperscript{44} In the United Kingdom and Europe, mass DNA screenings in small towns have been undertaken. The strategy has been employed in the United States as well. See E J Imwinkelried and D H Kaye ‘DNA Typing: Emerging or Neglected Issues’ 413 Wash L Rev 76 (2001).
million pairwise comparisons within racial groups, only two possible four-locus matches were found, and none were found for five or six loci.'

South Africa has, as the 1996 NRC report suggests, used convenience samples obtained from staff and from blood transfusion services. The report suggests that for the purpose of estimating allele frequencies, convenience sampling should give results comparable to random sampling, and it discusses procedures for estimating the random sampling error. In the earlier years of forensic DNA testing, the accused frequently contended that the size of the forensic databases was too small to give accurate estimates.

With regard to the size of the databases, Chakraborty states the following:

'Due to the fact that DNA typing is only an examination of a DNA sample's sequence and/or length at discrete locations, a match in DNA typing is always a statistical exercise. (Currently, time and expense limit an examination of an individual's entire genome, which would show unique identity.) In order to determine the probability that a particular genotype might occur at random in a population, population data must be gathered to make an estimate of the frequency of each possible allele and genotype. Usually a sample size of greater than 100 samples is sufficient to make reliable projections about a genotype's frequency in a larger population.'

The significance of a match or inclusion result

Once a match is declared, the next step is to compare the DNA profile to a population database. The question that arises here is how likely is the observation of such a match, if the accused is not in fact the person who left the crime scene sample? As remarked in United States v Yee

---

46 S v Motloutsi (Cape Provincial Division CC 48/95 unreported) S v Maqhina supra (n25).
47 R Chakraborty 'Sample size requirements for addressing the Population Genetics Issues of Forensic Use of DNA Typing' (1992) Human Biology 156-7. The 1996 NRC Report op cit (n20) 114, refers to 'at least several hundred persons,' but it has been suggested that relatively small databases, consisting of fifty or so individuals, allow statistically acceptable frequency estimates for the common alleles. A new, specially constructed database is likely to be small, but alleles can be assigned a minimum value, resulting in conservative genotype frequency estimates? Later, the NRC committee suggests that the uncertainty that arises '[i]f the database is small … can be addressed by providing confidence intervals on the estimates.' NRC II op cit (n20) 125.
49 49134 FRD 161 181 ND Ohio. In general there will be at least one sample from the crime scene and one or more from the accused. Each pair in a record will refer to a different locus (part of the DNA).
‘[w]ithout the probability assessment, the jury does not know whether the patterns are as common as two eyes, or as unique as the Mona Lisa.’

All genetic markers are found in the population with particular frequencies. Each allele of the markers used in DNA analysis also exhibits a particular population frequency. The frequency with which markers occur is extremely important in measuring the strength of a particular genetic type. If a type is found in half of the population, then the fact that the crime scene sample and the suspect share the type is not of particular significance in determining a link between the sample and the suspect. However, if the genetic type is found only in one out of 800 000 people then the fact that the sample and the suspect bear the same type, is of much greater significance and weight. In the case of the AmpFISTR Profiler Plus PCR amplification kit used by the SAPS Forensic Science Laboratory, ten different markers are tested. They are D3S1358; vWA; FGA; D8S1179; D2IS11; D18S51; D5S818; D13S317; D7S820 and Amelogenin. A typical DNA profile presented to a statistician in South Africa in 2004 will consist of a series of paired numbers for each DNA sample. \(^{50}\)

<table>
<thead>
<tr>
<th>DNA locus</th>
<th>D3S1358</th>
<th>vWA</th>
<th>FGA</th>
<th>D8S1179</th>
<th>AMEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>14, 15</td>
<td>15, 15</td>
<td>20, 21</td>
<td>13, 13</td>
<td>X:Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNA locus</th>
<th>D2IS11</th>
<th>D18S51</th>
<th>D13S317</th>
<th>D7S870</th>
<th>D5S818</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>30, 31.2</td>
<td>11, 12</td>
<td>8, 11</td>
<td>8, 9</td>
<td>10, 13</td>
</tr>
</tbody>
</table>

Chasely \(^{51}\) explains that the component of each pair represents the values of the two alleles (separate pieces received, one from mother, one from father) which the particular individual whose DNA profile has been measured has at each of these loci. The values will appear as two numbers separated by a non-numeric character such as a comma — see table above.

The more markers tested, and the more each corresponds with the suspect sample, the more powerful the results become. The most powerful way to express the overall strength of the match is to multiply the individual frequencies of the different markers. This is called the product rule. Much controversy has occurred concerning the methods used to estimate the probability of a multilocus DNA profile.

\(^{50}\) United States v Yee supra (n49) at 116–7.

\(^{51}\) Chasely op cit (n48).
The product rule debate

For purposes of completeness, an overview of the product rule debate is discussed here.

It has been argued that the probability calculation is only legitimate if the markers in question pass certain genetic and statistical tests. The probability model based on the product rule was widely challenged in the United States of America. The first requirement according to applicability of the product rule is that the population in question must approach the Hardy-Weinberg Equilibrium.

**Hardy-Weinberg Equilibrium**: a condition in which the allele frequencies within a large, random, intra-breeding population are unrelated to patterns of mating. In this condition, occurrence of alleles from each parent will be independent and have a joint frequency estimated by the product rule.

This means that the alleles at one locus show no a priori correlation with each other. If there is prior correlation it could increase the strength of genetic concordance artificially. The conditions are approached in large, random mating populations and in the absence of large changes due to migration, natural selection or gene mutation. The basic product rule estimates the frequency of genotypes in a population of individuals who choose their mates and reproduce independently of the alleles used to compare the samples. Although population geneticists describe this situation as random mating, these words are terms of art. Geneticists know that people do not choose their mates by a lottery, and they use ‘random mating’ to indicate that the choices are uncorrelated with the specific alleles that make up the genotypes in question. The second requirement for the application of the product rule was that the genotypes at the different genetic loci must lack correlations with each other in the population. This is known as linkage equilibrium.

**Linkage equilibrium**: a condition in which the occurrence of alleles at different loci is independent.

A population could deviate from Hardy-Weinberg equilibrium and/or linkage equilibrium if there is a large degree of substructure present due to distinct, but undetected sub-populations. Within the general population in Johannesburg, for example, some ethnic groups such as the Jews or the Portuguese may show a high rate of intra-ethnic marriage. Therefore, frequencies in such groups could deviate from those obtained by a sampling of the larger, mixed population.

---

52 LA Foreman et al ‘Interpreting DNA evidence: A review’ 2003 *International Statistical Review* 473 at 474. Issues that were raised in this debate included 1) the size and representativeness of the databases; 2) default band frequencies; 3) independence of results at the different loci; 4) measurement error and laboratory error; 5) dealing with relatives; and 6) mixed profiles from DNA that had come from more than one individual.

53 NRC II op cit (n20) 90.

54 The most prominent expression of this position is Lewontin and Hartl, ‘Population genetics in forensic DNA typing’ 254 *Science* (1991).
Extensive litigation and scientific commentary have considered whether the occurrences of alleles at each locus are independent events (Hardy-Weinberg equilibrium), and whether the loci are independent (linkage equilibrium frequencies do not follow the simple model of a homogeneous population mating without regard to the loci used in forensic DNA profiling). Lewontin and Hartl suggested that the major racial populations are composed of ethnic sub-populations whose members tend to mate among themselves.\(^{55}\)

For a brief time, a certain small number of scientists insisted that variations from one ethnic group to another within a race were larger than variations from one race to another.\(^{56}\) In the light of this literature,\(^{57}\) courts had grounds to conclude that the basic product rule, used with broad population frequencies, was not universally accepted for estimating profile frequencies within sub-populations. The 1996 NRC Report in the United States states that post-1992 research has rendered it unnecessary for the courts to use other calculation methods, but the basic product rule distinguishes between cases in which the suspect population is a broad racial population and those in which that population is a genetically distinct subgroup. In the former situation, Recommendation 4.1 endorses the basic product rule:

‘In general, the calculation of a profile frequency should be made with the product rule. If the race of the person who left the evidence-sample DNA is known, the database for the person’s race should be used; if the race is not known, calculations for all the racial groups to which possible suspects belong should be made’.

For example, the committee wrote,

‘If DNA is recovered from semen in a case in which a woman hitchhiker on an interstate highway has been raped by a white man, the product rule with the 2p rule can be used with VNTR data from a sample of whites to estimate the frequency of the profile among white males. If the race of the rapist were in doubt, the product rule could still be used and the results given for data on whites, blacks, Hispanics, and East Asians…(however) when there are partially isolated subgroups in a population, the situation is more com-

\(^{55}\) Compare Lewontin and Hartl op cit (n54) 1745 (‘there is, on average, one-third more genetic variation among Irish, Spanish, Italians, Slavs, Swedes, and other subpopulations than there is, on average, between Europeans, Asians, Africans, Amerindians, and Oceanians.’), with Richard C Lewontin Discussion, 9 Stat Sci 259, 260 (1994) (‘all parties agree that differentiation among [major ethnic groups] is as large, if not larger than, the difference among tribes and national groups [within major ethnic groups]’).

\(^{56}\) The literature on genetic differences across the globe is reviewed in for example, Devlin and Roeder op cit (n39) s 18 — 3.2.1, 725-8 (suggesting that this body of research indicates that the extent of the variation across subpopulations is relatively small).

\(^{57}\) Ibid.
plex; then a suitably altered model leads to slightly different estimates of the quantities that are multiplied together in the formula for the frequency of the profile in the population.’

Thus, the committee’s Recommendation 4.2 urges that:

‘[i]f the particular subpopulation from which the evidence sample came is known, the allele frequencies for the specific subgroup should be used as described in Recommendation 4.1. If allele frequencies for the subgroup are not available, although data for the full population are, then the calculations should use the population-structure equations 4.10 for each locus, and the resulting values should be multiplied.’

The committee recommended that the population-structure equations be used in special situations.

SANAS population-structure equations

SANAS, the South African National Accreditation System in its draft version 5 on Technical Guidelines for Forensic DNA Testing Laboratories set out specific guidelines in respect of population genetics. It states that relevant population(s) for which the frequency will be calculated should be identified. SANAS makes the following recommendations:

**Recommendation 1: Population groups:**

Forensic and Paternity laboratories have defined the following four major population groups: Asian, Caucasian, Black and Coloured. Calculations for all four major population groups to which possible suspects can belong are stated when reporting match probabilities. In parentage calculations only the one population group to which the involved parties belong is stated.

**Recommendation 2: Heterozygote and Homozygote profiles:**

The following Hardy-Weinberg formulae are applied to the following loci:

<table>
<thead>
<tr>
<th>Profile Type</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterozygote profiles</td>
<td>$2pq$</td>
</tr>
<tr>
<td>Homozygote profiles</td>
<td>$p^2$</td>
</tr>
</tbody>
</table>

Where p and q are the frequencies of the respective alleles.

---

58 Para 11.6.3 Technical Guidelines for Forensic DNA Testing Laboratories. The fact that it is in draft form does not mean that these principles are not applied by the SAPS-FL at present.
Recommendation 3: Multiple locus profiles:

Given the statistical independence of alleles at these loci in the NDSD, the frequency of a multi-locus genotype can be estimated by multiplying the genotype frequencies at each locus (product rule). A 95% Confidence Interval is applied to this answer and the upper limit (most conservative value) reported on.

Recommendation 4: Minimum allele frequencies:

Where alleles occur at a low frequency (less than five times in a population database) a minimum frequency of $\frac{5}{2N}$, where $N$ equals the number of individuals in the database, are assigned to these alleles.

Recommendation 5: Biological relationships, where requested:

If the possible contributors of the evidence sample include relatives of the suspect, DNA profiles of those relatives should be obtained. If these profiles cannot be obtained, the probability of finding the evidence profile in those relatives should only be calculated if requested by the court.

Measures of probative value

Random match probability

As seen above, the multiplication or product rule is used to compute a random match probability. The random match probability compares the accused to an arbitrarily chosen member of the defined population.\(^9\)

The source probability

Koehler\(^0\) indicates that the random match probability must be distinguished from the source probability. To determine that probability, the expert has to specify the size of the relevant population.

Likelihood ratios

To choose between two competing hypotheses, one can compare how probable the evidence is under each hypothesis. This approach

---

\(^9\) P Donnelly and RD Friedman ‘DNA database searches and the legal, consumption of scientific evidence’ 97 1999 Mich L Rev 931 at 938.

\(^0\) J Koehler ‘DNA matches and statistics: Important questions, surprising answers’ 76 1993 Judicature 222 at 224-5.
is completely acceptable provided that the DNA profile evidence is not presented in a format which mathematically combines with prior probabilities.

**Posterior probabilities or guilt probability**

The likelihood ratio expresses the relative strength of an hypothesis, but the judge or magistrate ultimately must assess a different type of quantity — the probability of the hypothesis itself. An elementary rule of probability theory known as Bayes' theorem yields this probability. The theorem states that the odds in light of the data (here, the observed profiles) are the odds as they were known prior to receiving the data times the likelihood ratio: posterior odds = likelihood ratio x prior odds. For example, if the relevant match probability were 1/100,000, and if the chance that the laboratory would report a match between samples from the same source were 0.99, then the likelihood ratio would be 99,000, and the bench could be told how the DNA evidence raises various prior probabilities that the defendant's DNA is in the evidence sample. It would be appropriate to explain that these calculations rest on many premises, including the premise that the genotypes have been correctly determined.

**Interpretation of DNA mixtures**

Forensic evidentiary samples often contain DNA from multiple contributors and therefore present a challenge for the DNA expert. A sample that contains DNA from two or more individuals is referred to as a mixture. A single person would usually contribute at the most two

---

61 Odds and probabilities are two ways to express chances quantitatively. If the probability of an event is P, the odds are P/(1-P). If the odds are 0, the probability is 0/0 + 1. For instance, if the probability of rain is 2/3, the odds of rain are 2 to 1: (2/3) / (1-2/3) = (2/3) / (1/3) = 2. If the odds of rain are 2 to 1, then the probability is 2/(2 + 1) = 2/3.

62 By 'relevant match probability,' we mean the probability of a match given a specified type of kinship or the probability of a random match in the relevant suspect population. For relatives more distantly related than second cousins, the probability of a chance match is nearly as small as for persons of the same subpopulation. Devlin and Roeder op cit (n 41) s 18-5.3.1.5 at 724.

63 For further discussion of how Bayes' rule might be used in court with DNA evidence, see, for example, NRC II op cit (n20) 201-3.

64 See Richard Lempert 'The honest scientist's guide to DNA evidence' 96 *Genetica* 119 (1995). If the decision-maker(s) accepted these premises and also decided to accept the hypothesis of identity over those of kinship and coincidence, it still would be open to the accused to offer explanations of how the forensic samples came to include his/her DNA even though he/she is innocent.

65 C Ladd, HC Lee, N Yang and FR Bieber 'Interpretation of complex forensic DNA mixtures' 2001 42(3) *Croatian Medical Journal* 244.
alleles for each locus. If there are more than two peaks visible at any locus, there is a strong possibility that the sample is a mixture. If three or more alleles show up at any locus, it may be difficult to determine whether the sample originated from two, three or even more individuals as various contributors may share many alleles.\footnote{WC Thompson, S Ford, T Doom, M Raymer and D Krane 'Evaluating forensic DNA evidence: essential elements of a competent defense review' April 2003 http://biofo-rensics.com/articles/champion1/champion1.html, accessed 17/01/2005.}

**Forensic DNA databank**

It was not until the introduction of the PCR amplification technology coupled to the analysis of short tandem repeats that a sufficiently sensitive and robust system was available for the formation of efficient and effective DNA databanks. A national databank can be extremely effective in the detection of perpetrators of crimes. The value of using DNA as source material in the investigation of violent and sexual crime is based on three premises according to Rudin and Inman:\footnote{N Rudin and K Inman *An Introduction to Forensic DNA Analysis* 2ed (2001) 157.}

- the tendency of violent criminals to repeat offending;
- violent and sexual crimes frequently result in biological evidence being shed by the perpetrator at the crime scene or in or on the victim; and
- the high discriminatory power of DNA to discriminate between individuals result in a very strong inference that the person, matched by the databank to the evidence left at the scene, is the source of the sample.

A DNA database including profiles of convicted offenders holds much promise for the resolution of violent crime cases, especially sexual offence cases where rates of recidivism are known to be exceptionally high. Forensic DNA databanks have been established in the United Kingdom, the United States, Canada, Australia and New Zealand. The United Kingdom currently has the largest forensic DNA databank in the world. The profiles created from these samples are stored on a national database. By June 2000 the database had linked evidence found at crime scenes to 75 000 offenders and had excluded more than 51 000 suspects from criminal investigation. The United States also has a national DNA databank system (CODIS). The CODIS system has only three indices of profile: convicted offenders (this is mostly individuals convicted of sexual offences and violent felonies), unknown suspects and an anonymous population used for statistical analysis.

There are a number of concerns about the use of DNA forensic testing and profiling. Generally, these concerns relate to consent and collection; the chain of custody and contamination; storage and destruction of DNA profiles and systems; privacy; and future expansion.
A South African DNA databank which includes the profiles of convicted offenders can contribute greatly to the administration of justice and the safety of South Africans by ensuring that those who commit serious crimes are identified more quickly, while innocent people are eliminated from suspicion. In order to reduce violent crime in South Africa, it will be imperative that research should be undertaken to draft legislation that balances the rights of convicted offenders against the rights of society to a safe environment and effective investigation and prosecution of violent crime through the aid of a DNA databank.

To make any databank effective, specific legislation should determine the powers and circumstances under which arrested and convicted individuals can be compelled to provide DNA samples for the purpose of future identification. Such legislation will need to be specific in the storage and retrieval provisions, and provide for the destruction of both the physical sample and the recorded data if that is to become policy. This legislation cannot be drafted without thorough comparative legal research. It is proposed that the legislation and practical implementation of DNA databanks in the United Kingdom, the USA, Canada, Australia, New Zealand and the European Union be examined. This research should be done as a matter of urgency. The SAPS Forensic Laboratory has the technology to cope with the profiling of DNA databank samples. Currently the equipment is underutilised.

**Conclusion**

DNA evidence has been the subject of extensive scrutiny by lawyers, judges, and the scientific community. Debate lingers over the safeguards that should be required in testing samples and in presenting the evidence in court. Moreover, there are many types of DNA analysis, and still more are being developed. New problems arise as advancing...
methods of analysis and novel applications of established methods are introduced.

As indicated above the basic theory and most of the laboratory techniques of DNA profiling are so widely accepted in the scientific world that disputed issues involve features unique to their forensic applications or matters of laboratory technique. In fifteen years DNA typing has made the transition from a novel set of methods for identification to a relatively mature and well studied forensic technology. However, one should not lump all forms of DNA identification together. New techniques and applications continue to emerge. Therefore, an article such as this cannot contain the most up-to-date information up at publication, but can only provide the solid background needed for lawyers to understand some of the most important legal and technical issues concerning DNA in the courts.

At the Spring 2004 meeting the President of American Prosecutors’ Research Institute, Newman Flanagan remarked: ‘The need to continue training criminal justice and forensic professionals about this revolutionary technology is indisputable.’

Glossary of Terms

Adenine (A). One of the four bases, or nucleotides, that make up the DNA molecule. Adenine only binds to thymine.

Allele. Alternative form of a genetic locus (e.g. at a locus for eye colour there might be alleles resulting in blue or brown eyes); alleles are inherited separately from each parent.

Amplification. Increasing the number of copies of a DNA region, usually by PCR.

Base. Component part of DNA nucleotides. Two of the DNA bases are pyrimidine in nature (cytosine and thymine), and the other two are purine (adenine and guanine). See also Adenine, Base Pair, Cytosine, Guanine, Nucleotide, Thymine.

Base pair. Two complementary nucleotides (A & T; C & G) held together by a weak hydrogen bond. A series of base pairs form nucleotides. See also Nucleotide.

Capillary electrophoresis. DNA samples are placed in a small, thin (capillary) tube filled with a gel or polymer. When the capillary is subjected to a high voltage current the DNA fragments migrate through the tube.

Cell. Basic units of living organisms which can be either unicellular or multicellular. An animal cell contains a nucleus, cytoplasm, mitochondria, and other organelles. Cells self-replicate through a process of cell division that includes copying all of its contents and then dividing in half.

Chromosome. Threadlike structures housed in the nucleus of cells on which genes are arranged in linear order. A full complement of chromosomes is 46 — 22 pairs of autosomes and two sex chromosomes.

Cytosine. One of the four bases that are found in nucleotides — the subunit of DNA. Cytosine, abbreviated ‘C’, binds only to Guanine. See also Base, Nucleotide, Guanine.

Degradation. The breaking down of DNA by chemical or physical means.

Denaturing. A process by which the hydrogen bonds on the original double-stranded DNA are broken or 'unzipped' leaving a single strand of DNA whose bases are available for hydrogen bonding.

Deoxyribonucleic acid (DNA). Genetic material present in the nucleus of a cell. This molecule contains all of the information necessary to code for all living things. Half of the material is inherited from each biological parent. DNA is organized into a double helix composed of two complementary chains of paired nucleotides. See also Cell, Double Helix, Nucleotide, Nucleus.

DNA sequence. The relative order of base pairs, whether in a fragment of DNA, a gene, a chromosome, or an entire genome.

Double helix. The shape that two paired strands of DNA assume when bonded together. A double helix is visually described as a twisting ladder.

Electrophoresis. The technique for separating large molecules by placing them in a medium (usually a gel) and applying an electric current. Molecules travel through the medium at different rates depending on their size. See also Capillary Electrophoresis, Gel Electrophoresis.

Enzyme. A protein that is capable of speeding up, but not changing the nature of a specific chemical reaction; a biological catalyst. See also Restriction Enzyme.

Gel. A semisolid medium used to separate molecules by electrophoresis. Forensic analysis usually utilizes an agarose or acrylamide gel to separate DNA molecules. See also Electrophoresis.

Gene. The fundamental unit of heredity. A gene is an ordered sequence of nucleotides located at a particular position on a particular chromosome. See Allele, Chromosome, Nucleotide.

Genotype. The genetic constitution of an organism, as distinct from its expressed features or phenotype. See also Phenotype.

Guanine. One of the four bases that are found in nucleotides — the subunit of DNA. Guanine, abbreviated ‘G’ binds only to Cytosine. See also Base, Nucleotide, Cytosine.

Hardy-Weinberg equilibrium — Refers to a population with random mating. In a human population, Hardy-Weinberg equilibrium results in independent association, a condition required in order to apply the product rule. See also Allele, Independent Association, Population, Product Rule.

Heterozygous. Having different alleles at a particular locus. See also Allele, Locus.
Homozygous. Having the same allele at a particular locus. See also Allele, Locus.

Hybridization. The process of pairing a single strand of DNA with its complementary strand by matching base pairs, usually with the assistance of a primer. See also Base Pair, Primer.

Linkage equilibrium. When all possible genotypes of a locus appear in a population with equal frequency.

Locus (Loci) -s. Locus, pl. Loci. The physical location of a gene on a chromosome. Any one of the possible alleles for a gene may be present at the gene's locus or along the genes' loci. See also Allele, Chromosome, Gene.

Marker. A gene of known location on a chromosome and phenotype that is used as a point of reference in the mapping of other loci.

Multiplexing. A test kit for analyzing several loci at once.

Nucleotide. A component part of DNA consisting of a base, a phosphate molecule, and a sugar molecule. Nucleotides are the raw building blocks of DNA. Nucleotides are paired according to the particular base and then linked to form alleles. See also Base, Base Pair.

Nucleus. A compartment within a cell that houses the chromosomes. The nucleus is separated from the cytoplasm and other organelles in the cell by the nuclear envelope. See also Cell, Chromosome.

Nucleolus. A small round body of protein in a cell nucleus that contain RNA and is involved in protein synthesis.

Polymerase. In DNA typing procedures, an enzyme that initiates the synthesis of double-stranded DNA. See also Enzyme.

Polymerase chain reaction (PCR). A process for amplifying (copying) DNA. Two primers target a particular DNA sequence (one primer for each complementary strand of DNA) to be amplified. In a series of cycles with varying temperatures, the DNA strand is denatured and copied with the help of a polymerase enzyme. Since each copy is denatured and copied in subsequent cycles, the DNA is amplified exponentially. See also Amplification, Denaturation, Enzyme, Polymerase.

Proficiency testing. A test to evaluate the competence of technicians and the quality of performance of a laboratory. Testing can be open or blind (depending on whether the person being tested is aware that the sample is part of a test) and internal or external (depending on whether the test is administered by the laboratory itself or an outside agency).

Product rule. When two or more loci are tested, the allele frequency at each locus is multiplied in order to estimate the overall frequency of that person's genetic profile. This formula assumes both linkage equilibrium and independent association. See also Locus, Independent Association.

Random match probability. The probability that the DNA in a random sample from the population will have the same profile as the DNA in the evidence sample.
Restriction enzyme. An enzyme that recognizes a specific series of nucleotides and cuts a DNA molecule wherever the series appears. See also Enzyme.

SANAS. The South African National Accreditation System.

Sequencing. Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Short tandem repeat (STR). Small regions of the DNA that contains short segments (usually 2, 3, 4, or 5 bases long) repeated several times in tandem (side-by-side).

Thymine. One of the four bases that are found in nucleotides — the subunit of DNA. Thymine, abbreviated ‘T’, binds only to Adenine. See also Base, Nucleotide, Adenine.

TWGDAM. Technical Working Group on DNA Analysis Methods. An organization made up largely of individuals from the FBI and public crime laboratories that recommend guidelines for DNA identification testing.

Validation. A process for the scientific community at large to properly assess whether a particular procedure can reliably obtain a desired result, determine the conditions under which such results can be obtained, and determine the limitations of the procedure.

X Chromosome. A sex chromosome present twice in female cells, and once in male cells. See also Autosome, Cell, Chromosome.

Y Chromosome. A sex chromosome present once in male cells, and transmitted directly from a father to all his sons. See also Autosome, Cell, Chromosome.